UNIVERSITÉ DE GRENOBLE

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Physique pour les Sciences du Vivant**

Arrêté ministériel : 7 août 2006

Présentée par

Romain TALON

Thèse dirigée par **Richard KAHN** et
codirigée par **Eric GIRARD**

préparée au sein de l'**Institut de Biologie Structurale**
Groupe Extremophiles and Large Macromolecular Assemblies
dans **l'École Doctorale de Physique**

DEVELOPPEMENT DE NOUVEAUX OUTILS POUR LA DETERMINATION DE LA STRUCTURE DE MACROMOLECULES BIOLOGIQUES PAR DIFFRACTION : APPLICATION AUX PROTEINES MEMBRANAIRES ET AUX GRANDS COMPLEXES PROTEIQUES.

Thèse soutenue le **6 juin 2012**
devant le jury composé de :

M. Vincent VILLERET
Directeur de Recherche, CNRS de Villeneuve d'Ascq, Rapporteur
M. Daniel PICOT
Directeur de Recherche, Université Paris Diderot de Paris, Rapporteur
M. Giuseppe ZACCAI
Directeur de Recherche, Institut Laue-Langevin de Grenoble, Membre
M. Andrew THOMPSON
Responsable de ligne de lumière, Synchrotron SOLEIL de Gif-sur-Yvette, Membre
M. Olivier MAURY
Directeur de Recherche de l'ENS de Lyon, Membre

A Richard Kahn,

feu mon directeur de thèse

Ont été imprimés 24 exemplaires de cette thèse,

10 au format A4 dont 3 uniques,

14 au format A5 dont 12 uniques.

Achevé d'imprimer en juin 2013

Notes

Le terme "**complexe**" désigne l'ensemble que constitue l'ion lanthanide (Gd^{3+}, Eu^{3+}...) avec le ligand. Le "**ligand**" étant la molécule chimique qui chélate l'ion lanthanide. Le mot "**dérivé**" caractérise l'ensemble que forme la protéine liée au complexe. Je l'emploierai notamment pour désigner les cristaux de protéines ayant été mis en présence des complexes de lanthanide, par opposition aux cristaux dits "**natifs**", uniquement constitués de protéines.

Afin de ne froisser aucun chimiste, je précise que, dans le cadre de cette thèse, le terme "**supramoléculaire**" désigne, pour un complexe de lanthanide, la propriété d'établir une ou plusieurs liaisons **avec au moins deux molécules de protéine distinctes** du cristal.

Pour la désignation des cartes de densité électronique, F_o et F_c correspondent respectivement aux modules des facteurs de structure **o**bservés et **c**alculés.

Les données biochimiques théoriques mentionnées dans ce manuscrit, telles que les points isoélectriques et les coefficients d'extinction molaire, ont été déduites de la séquence primaire à l'aide du programme "ProtParam" disponible sur le serveur d'outil bioinformatique ExPASy (*http ://web.expasy.org/protparam/*).

Les figures de modèles de macromolécules biologiques et de cartes de densités électroniques qui ornent ce manuscrit ont été réalisées à l'aide des logiciels **Coot**, version 0.6.1 (Emsley *et al.*, 2010) et **Pymol** version 1.31 (Schrodinger, 2010).

Les figures qui représentent les formules chimiques relatives aux différents composés ont été réalisées à l'aide du logiciel Chemdraw® Ultra de la société CambridgeSoft dans sa version 11.0 (Cousins, 2011).

Les surfaces électrostatiques présentées dans ce manuscrit ont été calculées à partir du modèle avec le **module APBS Tools2** pour **Pymol** dans sa version 2.1 (Baker *et al.*, 2001). Pour les calculs de Poisson-Boltzmann des surfaces électrostatiques, le module fait appel au programme PDB2PQR (Dolinsky *et al.*, 2004).

Glossaire

[P] Concentration en **p**rotéine.

(P) Volume de solution concentrée en **p**rotéine prélevé.

[C] Concentration en **c**omplexe

(C) Volume de solution concentrée en **c**omplexe prélevé

(R) Volume de solution prélevé dans un **r**éservoir de plaque de cristallisation.

APS **P**er**s**ulfate d'**a**mmonium

ß-DDM n-**d**o**d**écyl beta-D-**m**altoside

CHM 6-**C**yclo**h**exyl beta-D-**m**altoside

DMSO **Dim**éthyl**s**ulf**o**xyde

DO **D**ensité **o**ptique

EDTA Acide **é**thylène **d**iamine **t**étr**a**cétique

FPLC **C**hromatographie **r**apide des **p**rotéines en phase liquide

HCl Acide chloridrique

HEPES Acide 4-(2-hydroxyéthyl)-1-pipérazine éthane sulfonique

HPLC **C**hromatographie en phase **l**iquide à **h**aute **p**erformance

IPTG **I**so**p**ropyl ß-D-**t**hio**g**alactopyranoside

IRM **I**magerie par **r**ésonance **m**agnétique

MAD **M**ultiple-wavelength **a**nomalous **d**iffraction

$MgCl_2$ Chlorure de magnésium

MPD 2-**M**ethyl-2,4-**p**entane**d**iol

NaCl Chlorure de sodium

NAD **N**icotinamide **a**denine **d**inucleotide

PCR **R**éaction en **c**haîne par **p**olymérase

PEG MME : **P**oly**e**thylène **g**lycol **m**ono**m**éthyl **e**ther

PMSF **P**henyl**m**ethane**s**ulfonyl**f**luoride

RMSD Ecart quadratique moyen

SAD **S**ingle-wavelength **a**nomalous **d**iffraction

SDS **D**o**d**écylsulfate de **s**odium

SIRAS **S**ingle **i**somorphous **r**eplacement with **a**nomalous **s**cattering

TMAO **T**ri**m**ethyl**a**mine *N*-**o**xide

Tris 2-amino-2-hydroxyméthyl-1,3-propanediol

UV **U**ltra**v**iolet.

Index

AfUox, 83, 85, 94, 294
AfUox, cristallisation, 112
AvEGFP, 83, 85, 90
AvEGFP, biologie moléculaire, 294
AvEGFP, cristallisation, 111
AvEGFP, production, 296
AvEGFP, purification, 299

Biacore, 309

CaMalDH, 83, 85, 93, 113, 224, 240, 243
CaMalDH, cristallisation, 114, 214
CmCH34CnrXs, 83, 85, 92, 302
Combinatoire, 168

EcAcrB, 83, 85, 89, 270, 273
EcAcrB, production, 302
EcAcrB, purification, 303

HEWL, 99, 116, 119, 125, 128, 148, 149, 152, 182, 192, 287, 306
HEWL, cristallisation, 107
HvMalDH, 114, 231
HvMalDH, cristallisation, 214

LnCTDPA, 146, 167
LnDO3A, 67, 69, 227, 229, 284
LnDPA, 72, 80, 192, 198, 285
LnDTPA-BMA, 67, 227, 285
LnDTTA-BPA, 271
LnEATDPA, 146
LnHECTDPA, 146, 158
LnHETDPA, 146, 147
LnHMTDPA, 80, 100, 192, 284
LnHPDO3A, 67, 68, 216, 226, 285
LnTNTPA, 80, 180, 192
Luminescence, 65, 73, 103, 108, 110

MAD, 36

PhTET1-12s, 115, 198
PhTET1-12s, cristallisation, 198
PhTET2-12s, 83, 85, 95
Protocole d'affinement, 55
Protocole d'intégration des données, 53
Protocole MR, 29
Protocole phasage, 38
Protocole ShelxD, 33
PtMalDH, 235, 286
PtMalDH, cristallisation, 213

RMN, 314

SAD, 36
Seuils d'absorbtion, 46
SIRAS, 38
Supramoléculaire, 7, 75, 82, 97, 99, 192, 199, 201

TdTHAU, 118, 124, 127, 142, 148, 151, 183, 192, 306
TdTHAU, cristallisation, 109

Table des matières

Introduction générale **20**

I Du cristal biologique à la structure de la macromolécule : l'intérêt des complexes de lanthanide en biocristallographie **25**

1 Du cristal à la structure de la macromolécule biologique **27**

1.1 La diffraction par un cristal et le problème des phases 27

1.2 Les différentes méthodes de phasage 29

1.2.1 Le remplacement moléculaire 29

1.2.2 Les méthodes directes (méthodes de phasages *ab initio*) 31

1.2.3 Détermination de la sous-structure par les méthodes de Patterson 31

1.2.3.1 La fonction de Patterson 32

1.2.3.2 La fonction de Patterson anomale 32

1.2.3.3 La fonction de Patterson isomorphe 33

1.2.3.4 Choix du protocole de détermination des positions des diffuseurs anomaux 33

1.2.3.5 La synthèse de Fourier anomale 34

1.2.4 Un triste sire en ligne de mire, son seuil franchi, la triste Hélène de Troie devint folle : description des méthodes de phasage *de novo* 34

1.2.4.1 Les méthodes de remplacement isomorphe, SIR et MIR 34

1.2.4.2 Les méthodes exploitant la diffusion anomale, SAD et MAD 36

1.2.4.3 Les méthodes combinées SIRAS et MIRAS 38

1.2.4.4 Protocole retenu pour la détermination des phases par les méthodes SAD et MAD 38

1.2.4.5 Evaluation de la qualité des phases expérimentales dans *SHARP* 40

1.2.4.6 Evaluation de la qualité des phases expérimentales améliorées par *SOLOMON* 41

1.3 La diffusion anomale 41

1.3.1 Présentation du phénomène 41

1.3.2 Les raies blanches 45

1.4 La brisure de la loi de Friedel 47

1.5 Evaluation du signal anomal et de la qualité des données de diffraction 49

1.5.1 L'importance du taux d'occupation du diffuseur anomal choisi 49

1.5.2 L'importance de la résolution des données 49

1.5.3 L'importance du rapport signal sur bruit 50

1.5.4 L'importance de la redondance des données 50
1.5.5 La complétude 50
1.5.6 Les facteurs résiduels 50
1.5.6.1 Le facteur R_{merge} 51
1.5.6.2 Le facteur R_{pim} 51
1.5.6.3 Le facteur R_{ano} 51
1.5.7 Stratégie d'enregistrement des données de diffraction choisie au cours de ces travaux de thèse 51
1.5.8 Protocole d'intégration des données de diffraction retenu 53
1.5.8.1 Obtention des intensités 53
1.5.8.2 Calcul des amplitudes des facteurs de structure à partir des intensités . 55
1.5.8.3 Estimation de la proportion de solvant dans l'unité asymétrique 55
1.6 Affinement de la structure et interprétations 55
1.6.1 Les facteurs résiduels utilisés lors du processus d'affinement 55
1.6.1.1 Le facteur R_{work} 55
1.6.1.2 Le facteur R_{free} 56
1.6.2 Protocole d'affinement retenu pour nos travaux 56
1.6.3 Les cartes de densité électronique 58
1.6.4 Les facteurs de qualité d'affinement de structure 58
1.7 Des cristaux dérivés présentant un signal anomal pour un phasage *de novo* 59
1.7.1 Utilisation de protéines séléniées 59
1.7.2 Utilisation de la diffusion anomale des atomes intrinsèques 59
1.7.3 Utilisation de gaz nobles 60
1.7.4 Introduction d'atomes lourds par trempage et co-cristallisation 60
1.7.5 Protocole d'obtention de cristaux dérivés par trempage et par co-cristallisation avec les complexes de lanthanide 61
1.8 Utilisation des lanthanides en tant que diffuseurs anomaux 65
1.8.1 Les lanthanides, des diffuseurs anomaux sous-utilisés lorsqu'il s'agit de la préparation de cristaux dérivés 65
1.8.2 La luminescence des lanthanides 66

2 Utilisation des complexes de lanthanide en bio-cristallographie 67
2.1 Les complexes de lanthanide issus de l'imagerie médicale 67
2.1.1 Présentation des complexes IRM 67
2.1.2 Etat de nos connaissances sur le mode de fixation des complexes de lanthanide IRM 69
2.1.2.1 Le complexe LnHPDO3A, pour des interactions hydrophobes 69
2.1.2.2 Le complexe LnDO3A, pour les interactions ioniques 70
2.1.2.3 Les autres complexes IRM 71
2.1.3 La détection de la fixation des complexes IRM avant l'expérience de diffraction . 72
2.2 Un complexe de lanthanide luminescent : le tris-dipicolinate d'europium 73
2.2.1 Présentation 73

2.2.2 La détection de la fixation du complexe $[Eu(DPA)_3]^{3-}$ avant l'expérience de diffraction . . . 74
2.2.3 Etat de nos connaissances sur le mode de fixation du complexe $[Ln(DPA)_3]^{3-}$ . . 75
2.3 Protéines de structures inconnues déterminées jusqu'alors à l'aide des complexes de lanthanide . . . 76

II De nouveaux outils pour la détermination de la structure de macromolécules biologiques : la grande famille des complexes picolinates de lanthanide luminescents 79

3 Le tris-dipicolinate de lanthanide, agent de nucléation ? 83
3.1 Mise en place de l'expérience de co-cristallisation automatisée . . . 84
3.1.1 Choix des protéines modèles . . . 84
3.1.2 Choix des conditions de co-cristallisation . . . 84
3.1.3 Matériel & méthodes . . . 85
3.1.3.1 Composition des échantillons . . . 85
3.1.3.2 Observations des plaques . . . 86
3.1.3.3 Contrôle de la luminescence dans les gouttes . . . 86
3.1.4 La démarche retenue pour l'analyse statistique des expériences de co-cristallisation 87
3.2 Les observations issues de l'expérience de co-cristallisation automatisée avec le $[Eu(DPA)_3]^{3-}$ 89
3.2.1 Co-cristallisation de la protéine EcAcrB avec le complexe $[Eu(DPA)_3]^{3-}$ . . . 90
3.2.2 Co-cristallisation de la protéine AvEGFP avec le complexe $[Eu(DPA)_3]^{3-}$ . . . 91
3.2.3 Co-cristallisation de la protéine CmCH34CnrXs avec le complexe $[Eu(DPA)_3]^{3-}$. 93
3.2.4 Co-cristallisation de la protéine CaMalDH avec le complexe $[Eu(DPA)_3]^{3-}$ . . . 94
3.2.5 Co-cristallisation de la protéine AfUox avec le complexe $[Eu(DPA)_3]^{3-}$ . . . 95
3.2.6 Co-cristallisation de la protéine PhTET2-12s avec le complexe $[Eu(DPA)_3]^{3-}$ . . 96
3.3 Synthèse des observations et conclusion . . . 96

4 La famille des tris-triazoledipicolinates de lanthanide 101
4.1 Présentation du complexe luminescent $[Ln(HMTDPA)_3]^{3-}$ . . . 101
4.1.1 Caractéristiques chimiques du complexe . . . 102
4.1.2 Caractéristique spectroscopique : la luminescence dans le visible par effet d'antenne . . . 104
4.2 Préparation de cristaux dérivés avec le complexe $[Ln(HMTDPA)_3]^{3-}$ en vue d'une expérience de diffraction des rayons X . . . 107
4.2.1 Choix des protéines modèles et méthodes systématiquement utilisées . . . 107
4.2.2 Le lysozyme de blanc d'œuf de poule ("HEWL" pour Hen Egg White Lysozyme) 107
4.2.2.1 Cristallisation de la protéine native . . . 108
4.2.2.2 Obtention des cristaux dérivés . . . 108
4.2.3 La thaumatine de *Thaumatococcus daniellii* "TdTHAU" . . . 110
4.2.3.1 Cristallisation de la protéine native . . . 110
4.2.3.2 Obtention des cristaux dérivés . . . 111

4.2.4 La Protéine Fluorescente Verte de *Aequorea victoria* (“AvEGFP” pour *Aequorea victoria* Enhanced Green Fluorescent Protein) . . . 112
4.2.4.1 Cristallisation de la protéine native . . . 112
4.2.4.2 Obtention de cristaux dérivés . . . 112
4.2.5 L’urate oxydase de *Aspergillus flavus* “AfUox” . . . 113
4.2.5.1 Cristallisation de la protéine native . . . 113
4.2.5.2 Obtention des cristaux dérivés . . . 113
4.2.5.3 Recherche de conditions de co-cristallisation à l’aide du robot de cristallisation nanogoutte de l’EMBL . . . 113
4.2.6 La malate déshydrogénase de *Chloroflexus aurantiacus* “CaMalDH” . . . 114
4.2.6.1 Cristallisation de la protéine native . . . 114
4.2.6.2 Obtention des cristaux dérivés . . . 115
4.2.7 La malate déshydrogénase de *Haloferax volcanii* “HvMalDH” . . . 115
4.2.7.1 Cristallisation de la protéine native . . . 115
4.2.7.2 Obtention des cristaux dérivés . . . 115
4.2.8 L’aminopeptidase TET1-12s de *Pyrococcus horikoshii* “PhTET1-12s” . . . 116
4.2.8.1 Cristallisation de la protéine native . . . 116
4.2.8.2 Obtention des cristaux dérivés . . . 116
4.2.9 Constats expérimentaux concernant la préparation de cristaux dérivés avec le $[Ln(HMTDPA)_3]^{3-}$ . . . 116
4.3 Acquisition et traitement des données de diffraction . . . 117
4.3.1 Statistiques d’intégration des données pour les protéines présentant une fixation du complexe . . . 117
4.3.1.1 Intégration des données enregistrées sur des cristaux dérivés de lysozyme de blanc d’œuf de poule . . . 117
4.3.1.2 Intégration des données enregistrées sur des cristaux dérivés de thaumatine de *Thaumatococcus daniellii* . . . 119
4.4 Détermination des phases expérimentales . . . 120
4.4.1 Evaluation de la qualité des phases déterminées pour le lysozyme de blanc d’œuf de poule . . . 120
4.4.2 Evaluation de la qualité des phases déterminées pour la thaumatine de *Thaumatococcus daniellii* . . . 125
4.5 Affinement des structures et description du mode de fixation du $[Ln(HMTDPA)_3]^{3-}$ . . . 126
4.5.1 Statistiques des affinements des structures issues de cristaux dérivés . . . 126
4.5.1.1 Statistiques des affinements des structures du HEWL avec le $[Ln(HMTDPA)_3]^{3-}$ 126
4.5.1.2 Statistiques des affinements des structures de la TdTHAU avec le $[Ln(HMTDPA)_3]^{3-}$ 128
4.5.2 Description des modes de fixation du $[Ln(HMTDPA)_3]^{3-}$ . . . 129
4.5.2.1 Le $[Ln(HMTDPA)_3]^{3-}$ à la surface du lysozyme de blanc d’œuf de poule 129
4.5.2.2 La corrélation entre la méthode d’obtention de cristaux dérivés, le faciès cristallin et le mode de fixation . . . 139
4.5.2.3 Une fixation énantiospécifique du $[Ln(HMTDPA)_3]^{3-}$ . . . 142

4.5.2.4 Le $[Ln(HMTDPA)_3]^{3-}$ à la surface de la thaumatine de *Thaumatococcus daniellii* . . . 143
4.6 Le $[Ln(HMTDPA)_3]^{3-}$, un complexe modulable . . . 146
4.6.1 Cinq ligands dipicoliniques différents pour quatre complexes différents . . . 146
4.6.1.1 Récapitulatif des essais réalisés avec le complexe $[Eu(HETDPA)_3]^{3-}$ . . . 148
4.6.1.2 Récapitulatif des essais réalisés avec le complexe $Eu(HECTDPA)_3$ . . . 159
4.6.1.3 Récapitulatif des essais réalisés avec le complexe $[Eu(CTDPA)_3]^{6-}$ . . . 168
4.7 Le $[Ln(HMTDPA)_3]^{3-}$, un complexe modulable : la combinatoire . . . 169
4.7.1 Préparations des différentes combinatoires en fonction du but recherché . . . 170
4.7.2 Présentation des différentes combinatoires et résultats obtenus . . . 171
4.7.2.1 Les combinatoires avec mélange de complexes $[Ln(TDPA)_3]^x$ entiers . . . 171
4.7.2.2 La combinatoire 5, avec complexation de l'ion lanthanide avec deux ligands de la famille $[Ln(TDPA)_3]^x$ . . . 174
4.7.2.3 Les combinatoires avec mélange de complexes entiers appartenant à la grande famille des complexes picolinates . . . 176
4.7.2.4 La combinatoire 8, avec mélange de deux complexes dissemblables : le $[Eu(HMTDPA)_3]^{3-}$ et le GdHPDO3A . . . 177
4.8 Conclusions quant à l'utilisation des complexes de la famille des $[Ln(TDPA)_3]^x$ . . . 180

5 Un tris-dipicolinate cagé, le LnTNTPA 181
5.1 Description du complexe de lanthanide LnTNTPA . . . 181
5.2 Co-cristallisation du HEWL avec le complexe TbTNTPA : pas de fixation du TbTNTPA 183
5.3 Co-cristallisation de la TdTHAU avec le TbTNTPA : fixation du TbTNTPA . . . 184
5.3.1 Obtention de cristaux dérivés de TdTHAU avec le TbTNTPA . . . 184
5.3.2 Intégration des données de diffraction issues d'un cristal de TdTHAU obtenu par co-cristallisation avec le TbTNTPA . . . 185
5.3.3 Phasage expérimental de la TdTHAU avec le TbTNTPA . . . 185
5.3.4 Affinement du modèle et description du mode de fixation du TbTNTPA à la surface de la TdTHAU . . . 187
5.3.4.1 Description du site de fixation 1 (taux d'occupation : 20 %) . . . 188
5.3.4.2 Description du site de fixation 2 (taux d'occupation : 15 %) . . . 190
5.3.4.3 Le complexe TbTNTPA dans l'empilement cristallin d'un cristal dérivé de TdTHAU . . . 191
5.4 Résultats obtenus avec la CaMalDH . . . 191
5.5 Conclusion . . . 191

III Les complexes de lanthanide pour le phasage de protéines multimériques de grande taille 195

6 L'amino-peptidase dodécamérique TET1-12s de *Pyrococcus horikoshii* 199
6.1 L'amino-peptidase tétraédrique dodécamérique PhTET1-12s . . . 199
6.2 Obtentions de cristaux natifs et dérivés . . . 199

6.3 L'effet supramoléculaire et le pouvoir phasant du complexe $[Eu(DPA)_3]^{3-}$ 200
6.4 Conclusion 202

7 Les malate déshydrogénases issues d'organismes extrêmophiles 209
7.1 Les malate déshydrogénases pour une étude rationalisée du mode de fixation des complexes de lanthanide 210
7.1.1 Cristallisation des différentes malate déshydrogénases et obtention de cristaux dérivés 214
7.1.1.1 Cristallisation de la malate déshydrogénase de *Picrophilus torridus* . . . 214
7.1.1.2 Cristallisation de la malate déshydrogénase de *Haloferax volcanii* 215
7.1.1.3 Cristallisation de la malate déshydrogénase de *Chloroflexus aurantiacus* 215
7.1.2 L'étude de l'interaction des trois complexes IRM sur les trois MalDH 216
7.1.3 Etude de la fixation du GdHPDO3A sur trois MalDH 217
7.1.3.1 L'obtention de cristaux dérivés avec le complexe GdHPDO3A 217
7.1.3.2 Intégration des données de diffraction issues des cristaux dérivés de CaMalDH, PtMalDH et HvMalDH 217
7.1.3.3 Phasage des données de CaMalDH, PtMalDH et HvMalDH avec le GdHPDO3A 220
7.1.3.4 Affinement des structures de CaMalDH, PtMalDH et HvMalDH 223
7.1.3.5 Comparaison des surfaces électrostatiques, à région équivalente, des trois MalDH 225
7.1.4 Résumé des premières observations issues de l'étude de l'interaction des trois complexes IRM sur les trois MalDH 227
7.2 Les nouvelles structures de MalDH résolues par phasage *de novo* 231
7.2.1 HvMalDH, la malate déshydrogénase de *Haloferax volcanii* 232
7.2.1.1 Intégration des données en vue d'un phasage par la méthode SIRAS . . 232
7.2.1.2 Phasage par la méthode SIRAS 233
7.2.1.3 Affinement de la structure de HvMalDH 233
7.2.1.4 Une modélisation fiable des ions dans la structure de HvMalDH grâce à la diffusion anomale 234
7.2.2 PtMalDH, la malate déshydrogénase de *Picrophilus torridus* et la localisation de la boucle catalytique 236
7.2.2.1 Intégration des données de diffraction 236
7.2.2.2 Calcul des phases pour les deux types de données enregistrées 238
7.2.2.3 Affinement des structures de PtMalDH 238
7.2.2.4 Les deux conformations de la boucle catalytique de PtMalDH 240
7.2.3 CaMalDH, la malate déshydrogénase de Chloroflexus aurantiacus et la structure de l'eau dans son enveloppe de solvatation 241
7.2.3.1 Intégration des données utilisées pour le phasage par la méthode SIRAS 241
7.2.3.2 Phasage par la méthode SIRAS 242
7.2.3.3 Affinement de la structure de CaMalDH 242
7.2.3.4 Structuration de l'eau à la surface de CaMalDH et adaptation halophile 243
7.3 Conclusions et perspectives 256

8 AcrB, la protéine modèle membranaire putative 271
8.1 Les deux protagonistes de la courte histoire membranaire . . . 271
8.1.1 La protéine membranaire "EcAcrB" . . . 271
8.1.2 Le complexe de lanthanide DTTA-BPA . . . 272
8.2 Production, purification et cristallisation de EcAcrB . . . 274
8.2.1 Production & purification de EcAcrB . . . 274
8.2.2 Cristallisation de la protéine EcAcrB et cryo-protection des cristaux . . . 274
8.2.2.1 Cristallisation de la protéine EcAcrB . . . 274
8.2.2.2 Cryo-protection des cristaux avant l'étape de diffraction des rayons X . 274
8.3 Bref résumé des diverses tentatives d'amélioration . . . 275
8.3.1 Résumé des modifications apportées au protocole de purification d'EcAcrB . . . 275
8.3.2 Résumé des tentatives d'amélioration des conditions de cristallisation et délit de faciès cristallin . . . 276
8.3.3 Résumé des tentatives d'amélioration du protocole de cryo-protection des cristaux d'EcAcrB . . . 277
8.4 Conclusions quant à l'utilisation de la protéine modèle EcAcrB . . . 277

Discussion générale et perspectives 285

IV Annexes 291

9 Séparation des protéines par électrophorèse sur gel de polyacrylamide en présence de dodécylsulfate de sodium 293
9.1 Préparation des gels dénaturants de polyacrylamide . . . 293
9.2 Préparation du colorant au bleu de Coomassie . . . 294
9.3 Marqueurs de poids moléculaires utilisés . . . 294

10 Les protéines modèles 295
10.1 L'urate oxidase de *Aspergillus flavus* "AfUOx" . . . 295
10.2 L'enhanced Green Fluorescent Protein de *Aequorea victoria* "AvEGFP" . . . 295
10.2.1 Présentation de la protéine et données de biologie moléculaire . . . 295
10.2.2 Production et purification de l'AvEGFP . . . 297
10.2.2.1 Description du milieu de culture et de l'ensemble des tampons utilisés . 298
10.2.2.2 Préculture sur la nuit . . . 299
10.2.2.3 Culture bactérienne et production de la protéine . . . 299
10.2.2.4 Lyse cellulaire . . . 299
10.2.2.5 Chromatographie sur colonne d'affinité au nickel . . . 300
10.2.2.6 Colonne de désalinisation . . . 301
10.2.2.7 Clivage de l'étiquette histidine . . . 301
10.2.2.8 Chromatographie sur colonne échangeuse de cations . . . 301
10.2.2.9 Concentration de la protéine EGFP purifiée . . . 302

10.3 Forme soluble de la métalloprotéine membranaire CnrX du cpériplasme de la bactérie *Cupriavidus metallidurans* souche CH34 "CmCH34CnrXs" 303
10.4 L'Acriflavine resistance protein B "EcAcrB" 303
10.4.1 Production d'EcAcrB dans *E.coli* 303
10.4.2 Purification d'EcAcrB 304
10.4.2.1 Préparation et dosage des fractions membranaires de *E.coli* 304
10.4.2.2 Solubilisation des fractions de membranes cytoplasmiques 305
10.4.2.3 Purification de la protéine sur colonne d'affinité en résine de nickel . . . 305
10.5 Le lysozyme de blanc d'œuf de poule "HEWL" 307
10.6 La thaumatine de *Thaumatococcus daniellii* "TdTHAU" 307

11 Exploration d'autres méthodes pour révéler la fixation des complexes de lanthanide avant l'étape de diffraction 309
11.1 Détecter la fixation des complexes de lanthanide sur des macromolécules biologiques par Résonnance Plasmonique de Surface (Biacore®) 309
11.1.1 Principe théorique de la SPR et fonctionnement du Biacore® 310
11.1.2 Protocole général 311
11.1.3 Résumé très succinct des expériences menées 312
11.1.4 Résumé très succinct des résultats obtenus 313
11.1.5 Conclusion 314
11.2 Détecter la fixation des complexes de lanthanide sur des macromolécules biologiques par Résonance Magnétique Nucléaire 315
11.2.1 Brève description du principe théorique 315
11.2.2 Brève description du protocole retenu 315
11.2.3 Type de résultat observable par RMN 316
11.2.4 Conclusion 317

Remerciements 329

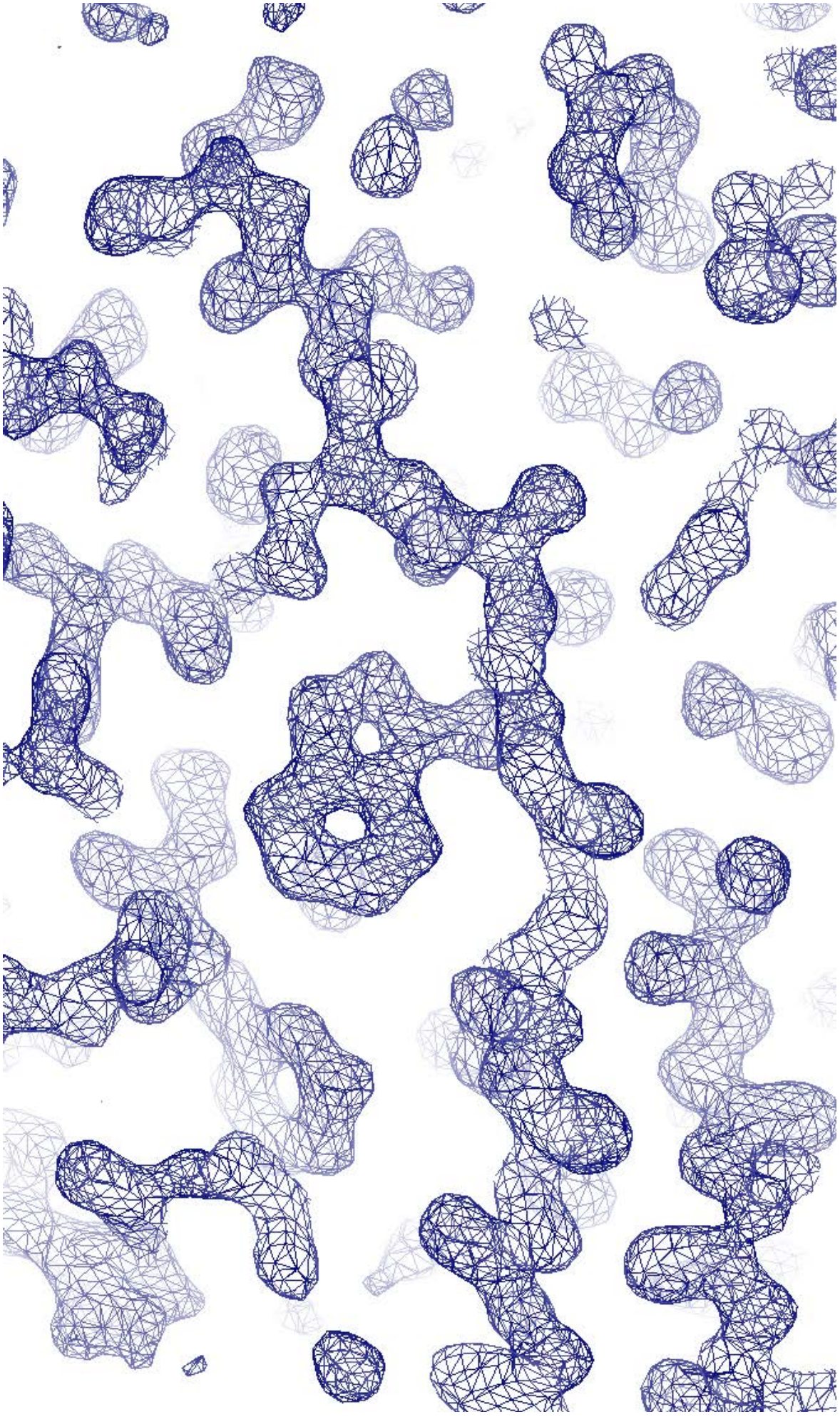


Figure 1 – Carte de densité électronique expérimentale d'un cristal dérivé de thaumatine avec le complexe d'europium $[Eu(HETDPA)_3]^{3-}$. Phasage MAD. 1,20 Å de résolution. Carte améliorée par aplatissement de solvant. Contours à 1,5 σ

Cette introduction présente les objectifs de mon travail de recherche. J'en profiterai aussi pour situer ce travail sans pour autant avoir la prétention de refaire un historique de la biologie structurale ou une énumération de toutes les techniques existantes permettant d'aboutir à la modélisation tridimensionnelle d'une macromolécule biologique.

Ce projet de recherche fondamentale est à l'interface de la physique, de la chimie et de la biologie. Il appartient au domaine de la biologie structurale, domaine que je pourrais définir par l'étude des structures de macromolécules biologiques pour la compréhension de leurs fonctions. La biologie structurale est ainsi complémentaire de la biochimie, approche qui permet de déduire la fonction de macromolécules biologiques par l'étude de leurs propriétés chimiques observables.

Or, l'observation de la structure d'une macromolécule biologique nécessite de descendre à l'échelle atomique, échelle qui a pour unité de longueur l'ångström : 1 Å = 1.10^{-10} m. "Observer" un objet à cette résolution atomique nécessite l'utilisation d'une lumière dont la longueur d'onde doit être du même ordre de grandeur, c'est à dire de l'ångström, gamme de longueurs à laquelle appartiennent les rayons X **[Figure 2]**. Ces rayons X peuvent être produits par des générateurs dédiés mais ils font surtout partie de la large gamme des longueurs d'onde couvertes par le rayonnement synchrotron (dont le spectre va de l'infrarouge aux rayons X durs). Particulièrement stable et brillant (car intense et focalisé), le faisceau synchrotron cohérent présente une faible divergence dans la gamme des rayons X durs. Cela permet, entre autres, la mise en œuvre d'expériences de diffraction des rayons X par un cristal de macromolécules biologiques. Pourquoi la nécessité d'un tel cristal ? D'autant plus qu'après les étapes biochimiques de production ou d'extraction, de purification et de concentration de la molécule souhaitée, la cristallisation de celle-ci constitue une étape problématique. Car, c'est le motif cristallin, soit la répétition périodique d'une maille élémentaire dans les trois directions de l'espace, qui permet le phénomène de diffraction : c'est à dire la diffusion élastique des rayons X, dans des directions spécifiques, par les électrons des atomes qui composent le cristal.

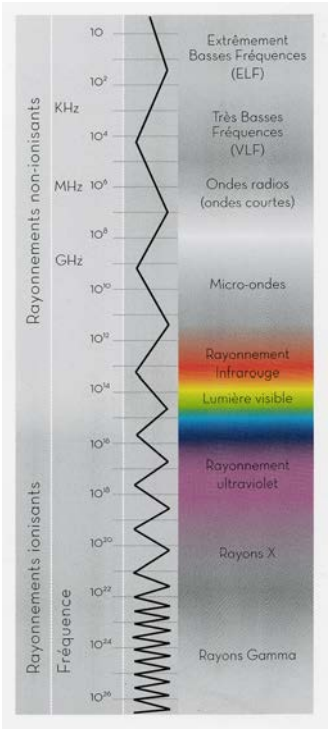


Figure 2 – Spectres d'ondes électromagnétiques (Feterman, 2010)

Les rayons X ainsi diffusés interfèrent de façon constructive ou destructive. Ce que nous mesurons sont les intensités des tâches de diffraction résultantes des rayons X lorsque ceux-ci, diffusés par le cristal-échantillon, ont traversé le détecteur. De la mesure de ces intensités, le cristallographe détermine l'information structurale à partir de laquelle un modèle moléculaire pourra être construit : la densité électronique du cristal **[Figure 1]**.

Travailler avec la diffraction des rayons X issus du rayonnement synchrotron suppose de travailler sans un couple de lentilles proprement dit qui, comme cela est le cas avec un microscope électronique, permettrait de recréer directement une image de l'échantillon. Pour un peu plus imager mon propos, je dirais qu'à la place de la première lentille d'un microscope, nous avons un détecteur enregistrant des ondes diffractées ; la seconde lentille, celle qui permet de reconstruire l'image réduite ou agrandie, est en fait recréée par calcul mathématique, la transformée de Fourier, mais elle ne permet pas de faire le

plan focal. Ainsi, une information cruciale pour obtenir la densité électronique n'est pas mesurée : le déphasage des ondes diffractées.

Mon travail de recherche se situe à ce niveau-là. Notre objectif est de développer l'utilisation de composés permettant de déterminer ces phases afin de résoudre expérimentalement la structure de protéines par diffraction des rayons X. Notre approche s'appuie sur l'introduction de métaux lourds dans les cristaux par le truchement des ligands chimiques chélatant l'atome lourd, en l'occurrence un lanthanide. Des méthodes de phasage seront alors mises en œuvre, en particulier celles exploitant la diffusion anomale de ces atomes lourds. La diffusion anomale est un phénomène caractéristique et mesurable qui apparait lorsque l'énergie des rayons X incidents est proche d'un des seuils d'absorption de l'atome en question. Ici réside le cœur de mon travail de recherche dont le but est double. D'une part, il est de parfaire notre compréhension des modes de fixation des complexes de lanthanide existants dans le but d'améliorer ces complexes ou d'en développer de nouveaux. D'autre part, il consiste à prouver qu'un très grand nombre d'acides aminés ne constituent pas un obstacle à l'utilisation de ces complexes de lanthanide pour l'obtention de phases précises, bien au contraire. Ce travail fait valoir ces outils comme alternatives à d'autres méthodes de phasage bien établies, comme l'incorporation de sélénométhionine ou de sélénocystéine dans les protéines.

Ce travail de recherche "méthodologique" fait suite à trois thèses au sein de l'équipe, initiées par celle de mon co-directeur de thèse, Eric Girard (1998 - 2001) et suivies de celles de Meike Stelter (2002 - 2005) et Guillaume Pompidor (2004 - 2007). Elles ont toutes démontré l'efficacité des complexes de lanthanide pour la résolution de structures de protéines. Il convient ici d'ajouter la thèse de Marie-Laure Chesne (1999 - 2002) qui mit à profit la diffusion anomale de l'uranium fixé aux protéines via un ligand chimique coloré. Cette thèse a ouvert la voie au développement de complexes dont les propriétés physico-chimiques les rendent détectables dans un cristal préalablement à l'expérience de diffraction, approche qui fut pleinement exploitée dans la thèse de Guillaume Pompidor.

Trois études constituent la première partie du travail présenté ici : elles s'articulent autour du complexe tris-dipicolinate de lanthanide, complexe dont le développement et l'étude ont fait l'objet de la thèse de Guillaume Pompidor et qui s'est révélé très prometteur. D'une part, sa fixation peut être aisément révélée par luminescence dans le visible et, d'autre part, il est susceptible de jouer le rôle d'agent de nucléation pour promouvoir l'apparition de cristaux dérivés en co-cristallisation. Ainsi, par une première étude statistique s'appuyant sur un large spectre de protéines, nous vérifierons si ses propriétés avantageuses font de lui un complexe si idéal. Nous présenterons donc les avantages et les inconvénients inhérents aux conditions d'utilisation de ce complexe. Au cours de ma recherche, cette étude s'est retrouvée étoffée par le développement de deux nouveaux complexes luminescents dérivés de ce tris-dipicolinate de lanthanide. Aussi, je présenterai donc une seconde étude menée sur ces nouveaux complexes, nommés tris-triazoledipicolinate de lanthanide. Enfin, je présenterai les travaux préliminaires réalisés sur un complexe cagé : le triazonanetripicolinate de lanthanide. Cette première partie est donc consacrée aux développements entrepris au sein de la famille des complexes luminescents basés sur le ligand picolinate.

Le second aspect de ce travail est à la fois méthodologique et appliqué. L'objectif est ici de démontrer que l'utilisation des complexes de lanthanide peut s'étendre aux protéines multimériques de grande taille, tant le signal anomal émis par les lanthanides est important et tant la fixation permise par les ligands est forte. Ceci sera mis en évidence dans le cas de protéines issues d'organismes extrêmophiles telles que

les malate déshydrogénases homotétramériques et l'aminopeptidase homododécamérique tétraédrique PhTET1-12s de *Pyroccocus horikoshii*. Enfin, dans cette seconde partie, nous poserons les bases d'une nouvelle approche pour approfondir notre connaissance des modes de fixation des complexes issus de l'imagerie médicale.

Première partie

Du cristal biologique à la structure de la macromolécule : l'intérêt des complexes de lanthanide en biocristallographie

Chapitre 1

Du cristal à la structure de la macromolécule biologique

Dans ce tout premier chapitre, nous allons exposer le problème des phases en bio-cristallographie et les méthodes expérimentales utilisées pour le résoudre. Le but de cette introduction n'est pas de détailler toutes les méthodes de détermination de phase existantes, mais plutôt de donner les outils nécessaires à la compréhension des expériences de diffraction décrites par la suite. Après une brève présentation des théories physiques exploitées lors de nos travaux de recherche, nous consacrerons la seconde partie de ce chapitre à la description des différentes étapes qui conduisent, depuis l'enregistrement des données, à l'obtention d'un modèle affiné. Comme une approche méthodologique requiert aussi une utilisation minutieuse des programmes de bio-cristallographie et une réelle reproductibilité, je profiterai de cette seconde partie pour consigner l'ensemble des protocoles et des méthodes que nous avons choisis pour mener à bien chacune de ces étapes.

1.1 La diffraction par un cristal et le problème des phases

Comme nous l'avons exposé au tout début du manuscrit, l'interaction entre le rayonnement X et la matière est la diffusion élastique par les électrons. Si cette matière est un cristal, de par le motif cristallin, c'est à dire la répétition périodique d'une maille élémentaire dans les trois directions de l'espace, l'onde est diffusée dans des directions spécifiques. On parle alors de diffraction. Les plans d'atomes (ou plans réticulaires) du cristal vont réfléchir les rayons avec formation d'interférences liées à la différence de marche des rayons X diffractés par deux plans voisins. Chaque fois que la différence de marche est égale à un multiple de la longueur d'onde, on observe une interférence constructive, c'est la loi de Bragg :

$$2dsin\theta = \mathrm{n}.\lambda$$

Avec :

d**,** distance inter-réticulaire (distance entre deux plans réticulaires)

θ**,** angle de Bragg (moitié de l'angle entre le faisceau incident et la direction du détecteur)

n**,** ordre de diffraction

$\boldsymbol{\lambda}$**,** longueur d'onde des rayons X

Chaque rayon X diffracté qui a traversé le plan vertical que constitue le détecteur génère une "tâche de diffraction" ou "réflexion". L'ensemble des tâches de diffraction provenant des rayons X diffractés ayant traversé le détecteur pendant le même intervalle de temps Δt est enregistré et constitue le "cliché de diffraction". Seules sont visibles sur les clichés de diffraction les réflexions résultantes des rayons X ayant interféré de façon constructive. La position de ces réflexions sur les clichés de diffraction forme le réseau réciproque (dans "l'espace réciproque", abstrait) et correspond à la transformée de Fourier du réseau cristallin (qui, lui, se trouve dans "l'espace réel"). On assigne à chaque réflexion un point particulier du réseau réciproque, à l'aide de trois indices h, k, l, nommés indices de Miller.

La mesure de l'intensité $I(hkl)$ de chacune de ces réflexions permet de déterminer l'amplitude de l'onde diffractée dans la direction correspondante, puisque l'intensité est proportionnelle au carré de l'amplitude. Or, pour calculer la densité électronique à partir de ces mesures, il faut connaître non seulement l'amplitude de l'ensemble des ondes diffractées mais aussi leur déphasage par rapport au rayonnement incident. Cela revient à connaître, pour la réflexion considérée, l'amplitude complexe de chacune des ondes diffractées. Cette amplitude complexe, transformée de Fourier de la densité électronique, est appelée facteur de structure, $F(hkl)$.

Pour les réflexions $(hkl) \equiv \mathbf{h}$, l'intensité $I(\mathbf{h})$ est donc proportionnelle au carré du module de facteur de structure $F(\mathbf{h})$:

$$I(\mathbf{h}) = F(\mathbf{h}) \cdot \overline{F}(\mathbf{h}) = |F(\mathbf{h})|^2$$

où $\overline{F}(\mathbf{h})$ est le complexe conjugué. Le facteur de structure $F(\mathbf{h})$ de l'atome j , de coordonnées réduites $\mathbf{x}_j \equiv (x_j, y_j, z_j)$ à l'intérieur de la maille, se décrit par :

$$F(\mathbf{h}) = \sum_j f_j \exp(i2\pi\mathbf{h} \cdot \mathbf{x}_j)$$

où j varie de manière à décrire l'ensemble des atomes de la maille cristalline. Et avec :

f_j = facteur de diffusion atomique de l'atome j , transformée de Fourier de la densité électronique de l'atome j.

Les facteurs de structure sont des nombres complexes avec une amplitude $|F(\mathbf{h})|$ et une phase $\phi(\mathbf{h})$ que l'on peut donc écrire :

$$F(\mathbf{h}) = |F(\mathbf{h})| \exp(i\phi(\mathbf{h}))$$

La densité électronique $\rho(\mathbf{r})$ se calcule par la transformée de Fourier inverse des facteurs de structure de chaque atome :

$$\rho(\mathbf{r}) = TF^{-1}(F(\mathbf{h})) = \frac{1}{V}\sum_{\mathbf{h}} F(\mathbf{h}) \exp(-i2\pi\mathbf{h} \cdot \mathbf{x}) = \frac{1}{V}\sum_{\mathbf{h}} |F(\mathbf{h})| \exp(i\phi(\mathbf{h})) \exp(-i2\pi\mathbf{h} \cdot \mathbf{x})$$

où V est le volume de la maille cristalline.

La mesure de l'intensité $I(\mathbf{h})$ ne peut fournir que le module du facteur de structure des atomes $|F(\mathbf{h})|$. Or, la phase $\phi(\mathbf{h})$ est indispensable pour le calcul de la synthèse de Fourier qui conduit à la densité

électronique $\rho(\mathbf{r})$: c'est ce que l'on nomme le "problème des phases". Une expérience de diffraction sur un cristal natif ne permet donc pas, à elle seule, la résolution de la structure.

1.2 Les différentes méthodes de phasage

Parmi les méthodes permettant de calculer ces phases, nous nous intéresserons plus particulièrement aux méthodes de phasage *de novo*.

1.2.1 Le remplacement moléculaire

Avant d'exposer les méthode de phasage *de novo*, il convient de présenter la méthode du remplacement moléculaire, méthode la plus couramment utilisée pour la résolution de structures de macromolécules biologiques (Rossmann, 1972; Taylor, 2010). Cette méthode ne fait pas partie des méthodes de phasage *de novo* car l'information de phase recherchée provient d'un modèle déjà existant et que l'on suppose proche de la structure à déterminer.

Le choix du modèle est orienté par le postulat qui stipule que l'homologie de séquences entre deux protéines implique très fréquemment une similarité structurale. Le modèle doit présenter une forte identité de séquences (au moins 30 %) avec celle de la protéine dont la structure est à déterminer. En outre, le modèle choisi peut correspondre à l'ensemble ou à une partie de la structure d'intérêt.

Une fois le modèle choisi, il faut alors rechercher la position et l'orientation de ce modèle dans la maille qui conduit au calcul des intensités qui correspondent aux intensités observées, tout en évitant le chevauchement de molécules symétriques. Les phases obtenues pour les modèles, positionnés et orientés, qui donnent les meilleurs accords sont appliquées aux données expérimentales. L'avantage de cette méthode est qu'elle ne nécessite que des données de diffraction d'un cristal natif. De plus, le nombre croissant de structures résolues enrichit le "catalogue" de modèles disponibles.

Cependant, cette méthode peut être mise en défaut dans deux occasions. D'un part, si la structure de la protéine à déterminer présente un nouveau repliement et/ou diffère significativement des modèles existants, le recours aux méthodes de phasage expérimental est nécessaire. D'autre part, la solution obtenue peut être biaisée par les informations apportées par le modèle choisi et dans ce cas, l'information introduite peut être difficile à supprimer ou à modifier au cours du processus d'affinement.

Protocole que nous avons utilisé en cas d'échec du phasage *de novo*

Au cours de ce travail de thèse, comme décrit par la suite, nous nous sommes parfois heurtés à des situations où la contribution anomale des atomes n'a pas toujours permis d'aboutir à la construction d'un modèle suite à un phasage expérimental. De plus, si la synthèse de Fourier anomale (*cf.* ci-après) présentait toutefois des pics dont la hauteur préfigurait un taux d'occupation suffisant pour modéliser une partie du complexe, nous avons eu alors recours au remplacement moléculaire. En guise de modèle, nous avons toujours choisi une structure de la même protéine déjà résolue par nos soins. Nous avons mis au point un protocole de phasage dont le but est de permettre de s'affranchir au maximum du biais que pourrait introduire le modèle choisi. Voici la description de ce protocole :

1. Construction d'un modèle protéique de départ épuré pour s'affranchir au maximum du biais-modèle :

   (a) Si la protéine à étudier est oligomérique, seul un monomère est conservé pour le modèle initial,

   (b) tous les atomes ne faisant pas partie de la protéine ainsi que ceux qui constituent toutes les chaines latérales des acides aminés de la protéine sont effacés. Le modèle réside alors en une seule chaine principale poly-alanine.

2. Obtention d'un jeu de phases par remplacement moléculaire à l'aide du modèle décrit en 1) et du programme *Phaser* de la *suite CCP4* :

   (a) Les paramètres utilisés ont été : le nombre de molécules présentes dans l'unité asymétrique et la recherche de la solution à la plus haute résolution à partir des amplitudes mesurées, tant dans le groupe d'espace indiqué que dans les groupes d'espace énantiomorphes (si il y en a). En effet, pour des cristaux de groupes d'espace centrosymétriques, la question de la main ne se pose pas : la configuration tridimensionnelle périodique des atomes et son inverse conduit au même profil de diffraction. Mais dans le cas de groupes d'espace non-centrosymétriques, cela conduit à une ambiguité de main (Blundell and Johnson, 1976).

3. Calcul d'une carte de densité électronique grâce aux phases alors obtenues et reconstruction d'un modèle :

   (a) Amélioration des phases calculées par *Phaser* à l'aide du programme *DM* de la *suite CCP4* (Cowtan and Main, 1996), en absence de tout modèle. Par modification de densité, le programme *DM* affine les phases calculés précédemment par le programme *Phaser* (*cf.* protocole utilisé pour les expériences SAD et MAD). Lors de cette étape, la "combinaison de phases" a toujours été désactivée afin que le calcul de la carte aplatie s'affranchisse un peu plus du biais qu'a pu introduire le modèle dans le calcul des phases initiales. Dans ce cas, la combinaison de ces phases initiales avec les nouvelles phases améliorées n'est plus utilisée et seules ces dernières sont utilisées.

   (b) Construction automatique de modèle par le programme *Buccaneer* (Cowtan, 2006) à l'aide des amplitudes expérimentales, des phases améliorées et de la séquence primaire de la protéine, sans cycle d'affinement entre les cycles de construction.

A l'issue de ce remplacement moléculaire, trois indicateurs statistiques sont calculés permettant de juger la qualité des solutions proposées : le gain du logarithme de maximum de vraisemblance global ("LLG") et les "Z-scores" des fonctions de rotation ("RFZ") et de translation ("TFZ") du modèle orienté.

La valeur du "LLG" indique à quel point les données calculées à partir du modèle orienté sont meilleures que celles qui auraient été prédites à partir d'un modèle atomique aléatoire (autrement dit, qu'un maximum de vraisemblance calculé à partir d'une distribution de Wilson). Si cette valeur est négative, cela signifie que notre modèle orienté est moins bien choisi qu'un modèle atomique aléatoire. Pour la solution choisie, la valeur du LLG doit donc être positive. Aussi permet-elle d'estimer la précision de la solution : plus elle est élevée, plus la solution est correcte. *Note : d'après le manuel d'utilisation, une valeur de LLG globale supérieure à 400 caractérise une solution non-équivoque.*

Les "Z-scores" résultant des fonctions de rotation ("RFZ") et de translation ("TFZ") sont calculés ainsi :

$$Z - score = \frac{LLG - |LLG\,d'un\,échantillon\,aléatoire\,d'orientations|}{RMSD^{|LLG|}_{LLG\,aléatoire}}$$

où $RMSD^{|LLG|}_{LLG\,aléatoire}$ est l'écart type calculé entre une distribution de valeurs de LLG aléatoires et la valeur moyenne mesurée du LLG.

En d'autres termes, la valeur de RFZ et de TFZ indique respectivement : le nombre d'écarts type supérieurs à la moyenne pour une valeur de LLG résultant de la rotation et le nombre d'écarts type supérieurs à la moyenne pour une valeur de LLG résultant de la translation du modèle. *Note : d'après le manuel d'utilisation, un Z-score supérieur à 5 prouve une orientation correcte du modèle.*

1.2.2 Les méthodes directes (méthodes de phasages *ab initio*)

Sous le terme de « méthodes directes », on regroupe un ensemble de méthodes qui s'attachent à résoudre le problème des phases en utilisant des relations entre phases qui découlent uniquement des intensités observées.

Ces relations mettent en oeuvre les facteurs de structure normalisés, $E(hkl)$. Ce sont des nombres complexes calculés en divisant les facteurs de structure $F(hkl)$ par la racine carrée de l'intensité moyenne, intensité qui serait obtenue à une résolution donnée si les atomes étaient distribués au hasard dans la maille. Les relations de phases les plus fortes sont obtenues entre les réflexions ayant les valeurs les plus élevées pour le module du facteur de structure normalisé, $|E(hkl)|$ (Bricogne, 1988). La valeur théorique maximale que peut prendre le module du facteur de structure normalisé est proportionnelle à la racine carrée du nombre d'atomes dans la maille. Cette valeur serait obtenue si les contributions au facteur de structure des atomes de la maille étaient toutes en phase pour la réflexion (hkl) considérée. La probabilité d'un tel événement décroît très rapidement avec le nombre d'atomes dans la maille. Aussi, les relations entre phases s'amenuisent lorsque le nombre d'atomes dans la maille augmente. C'est pour ces raisons que l'utilisation de ces méthodes est limitée à la détermination de structures de petites molécules.

Ces méthodes directes sont presque toujours utilisées pour résoudre la structure de petites molécules constituées de quelques centaines d'atomes. On peut aussi les exploiter pour déterminer la structure de macromolécules biologiques sans atome lourd mais cela reste limité à des structures constituées d'environ un millier atomes et pour lesquelles les données de diffraction enregistrées sont à très haute résolution ($d_{réso} < 1{,}2$ Å) (Usón and Sheldrick, 1999).

Utilisant les méthodes de phasage expérimental, nous utilisons les méthodes directes pour rechercher la position d'atomes lourds dans un cristal dérivé. Elles permettent ainsi de déterminer la sous-structure qui sert de référence pour résoudre la structure de la macromolécule d'intérêt par phasage expérimental.

1.2.3 Détermination de la sous-structure par les méthodes de Patterson

La détermination de la position des diffuseurs anomaux peut se faire soit par méthode directe soit par méthode de Patterson (Patterson, 1934). Celle que nous avons utilisée est la méthode de Patterson décrite ici.

1.2.3.1 La fonction de Patterson

La fonction de Patterson est définie comme le développement en série de Fourier des intensités mesurées expérimentalement $I = |F(\mathbf{h})|^2$.

$$P(\mathbf{u}) = \frac{1}{V} \sum_{\mathbf{h}} |F(\mathbf{h})|^2 \exp(-i2\pi\mathbf{h} \cdot \mathbf{u})$$

avec :

V**,** le volume de la maille cristalline.

$\mathbf{u} = u\mathbf{a} + v\mathbf{b} + w\mathbf{c}$**,** vecteur unitaire dans l'espace réciproque

En l'absence de diffusion anomale (par conséquent, la loi de Friedel est respectée), la fonction de Patterson devient la fonction d'autocorrélation de la densité électronique $\rho(\mathbf{r})$:

$$P(\mathbf{u}) = \rho(\mathbf{r}) * \rho(-\mathbf{r}) = \int_V \rho(\mathbf{r})\rho(\mathbf{u} + \mathbf{r}) d^3\mathbf{r}$$

Pour des densités électroniques décrites avec des atomes ponctuels, la densité s'écrit :

$$\rho(\mathbf{r}) = \sum_j Z_j \delta(\mathbf{r} - \mathbf{r}_j)$$

où Z_j est le nombre d'électrons de l'atome j de cordonnées réduites $\mathbf{r}_j$. Avec ce modèle, la fonction de Patterson s'écrit alors :

$$P(\mathbf{u}) = \int_V \sum_j Z_j \delta(\mathbf{r} - \mathbf{r}_j) \sum_{j'} Z_{j'} \delta(\mathbf{u} + \mathbf{r} - \mathbf{r}_{j'}) d^3\mathbf{r} = \sum_j \sum_{j'} Z_j Z_{j'} \int_V \delta(\mathbf{r} - \mathbf{r}_{j'}) \delta(\mathbf{u} + \mathbf{r} - \mathbf{r}_{j'}) d^3\mathbf{r}$$

$$P(\mathbf{u}) = \sum_j \sum_{j'} Z_j Z_{j'} \delta(\mathbf{u} - \mathbf{u}_{jj'})$$

Avec, $\mathbf{u}_{jj'} = \mathbf{r}_{j'} - \mathbf{r}_j$, les vecteurs interatomiques de la maille. Ainsi, la fonction de Patterson présente des pics de hauteur $Z_j Z_{j'}$ aux extrémités des vecteurs interatomiques.

1.2.3.2 La fonction de Patterson anomale

Lorsque les différences anomales sont significatives, on peut calculer la fonction de Patterson anomale $P_{ANO}(\mathbf{u})$avec les coefficients $D'_{ANO} = D_{ANO} \frac{^0f}{2f''}$ où $D_{ANO} = |F(+)| - |F(-)|$ est la différence de Bijvoet. Ainsi :

$$P_{ANO}(\mathbf{u}) = \frac{1}{V} \sum_{\mathbf{h}} |D'_{ANO}(\mathbf{h})|^2 \exp(-i2\pi\mathbf{h} \cdot \mathbf{u})$$

On peut montrer que $P_{ANO}(\mathbf{u})$correspond, au bruit près, à la fonction de Patterson avec pour coefficients les facteurs de structure des diffuseurs anomaux :

$$P_{ANO}(\mathbf{u}) \approx P_H(\mathbf{u}) = \frac{1}{V} \sum_{\mathbf{h}} |F_H(\mathbf{h})|^2 \exp(-i2\pi\mathbf{h} \cdot \mathbf{u})$$

Avec $F = F_P + F_H$ où :

F_P est le facteur de structure de la protéine native,

F_H est le facteur de structure des diffuseurs anomaux.

La fonction de Patterson anomale représente alors la carte des vecteurs interatomiques des atomes lourds.

1.2.3.3 La fonction de Patterson isomorphe

De la même manière, dans le cas d'un phasage par remplacement isomorphe, la fonction de Patterson ayant alors pour coefficients $D_{ISO} = |F_{PH}| - |F_P|$ correspond, au bruit près, à :

$$P_{ISO}(\mathbf{u}) \approx P_H(\mathbf{u}) = \frac{1}{V} \sum_{\mathbf{h}} |F_H(\mathbf{h})|^2 \exp(-i2\pi\mathbf{h} \cdot \mathbf{u})$$

Les fonctions de Patterson anomal et isomorphe présentent donc respectivement des pics à l'extrémité des vecteurs interatomiques des diffuseurs anomaux et des atomes lourds.

Dans le cas d'un cristal natif, compte tenu du grand nombre de pics interatomiques, la fonction de Patterson n'est en général pas interprétable. A l'inverse, la carte de Patterson anomale présente des pics beaucoup moins nombreux. La présence de pics marqués dans la carte de Patterson anomale indique la fixation de diffuseurs anomaux sur la protéine. A partir de cette carte, on peut trouver les positions des atomes lourds dans la maille. Les éléments de symétrie, imposés par le groupe d'espace, mènent à des concentrations de pics dans certaines régions de la carte de Patterson, appelées " sections de Harker ". Les pics trouvés dans ces sections indiquent les extrémités de vecteurs interatomiques d'atomes anomaux occupant des positions équivalentes. Les pics apparaissant en dehors des sections de Harker sont relatifs à des diffuseurs anomaux ne se trouvant pas dans des positions équivalentes, ce sont des pics croisés de différents sites. Connaissant les relations qui relient les positions équivalentes, les coordonnées x,y,z des diffuseurs anomaux peuvent être déduites des coordonnées u,v,w des pics des sections de Harker. La cohérence du résultat est vérifiée par inspection des pics croisés.

Alors que pour un faible nombre de sites anomaux on peut interpréter soi-même une carte de Patterson, on utilise généralement des programmes de détection des sites des atomes lourds.

1.2.3.4 Choix du protocole de détermination des positions des diffuseurs anomaux

Différents programmes permettent de rechercher les positions des diffuseurs anomaux dans un cristal à partir de données SIR, MIR, SAD, SIRAS ou MAD. Le programme *ShelxD* (Sheldrick, 2008) qui s'appuie, entre autre, sur les méthodes de déconvolution de Patterson est le programme le plus connu pour le phasage *de novo*. Il existe aussi le programme *Shake'n Bake* (Miller *et al.*, 2007) qui s'appuie sur des méthodes similaires.

Au cours de cette thèse, nous avons utilisé le programme *ShelxD* en complément du programme *HySS* (Hybrid Substructure Search) de la *suite Phenix* (Grosse-Kunstleve and Adams, 2003) pour la détermination des atomes lourds. Le programme *ShelxD* regroupe les deux méthodes (interprétation de la carte de Patterson dans l'espace réciproque puis utilisation des méthodes directes). Après identification

des sites, nous avons vérifié la correspondance entre les pics calculés et ceux observés dans la carte de Patterson anomale.

Paramètres du programme *HySS* de la *suite Phenix* et du programme *ShelxD* de la *suite CCP4* Que ce soit avec l'un ou l'autre des programmes, la recherche de de la position des lanthanides a toujours été réalisée en utilisant les intensités mises à l'échelle et leur sigma associé : $I(+)$, $SIGI(+)$, $I(-)$, $SIGI(-)$. En ce qui concerne le programme *ShelxD*, la recherche d'un nombre de sites a été réalisée sur plus de 500 itérations. Bien qu'il ait la même utilité que le programme *ShelxD*, j'ai davantage utilisé le programme *HySS*, (avec, comme indication, un nombre de sites attendus égal au nombre de sites identifiés et confirmés sur la carte de Patterson anomale) pour la création des fichiers d'entrée nécessaires au programme *ShelxD.*

1.2.3.5 La synthèse de Fourier anomale

Comme nous l'avons vu, la synthèse de Fourier permet de passer de l'espace réciproque vers l'espace réel et vice-versa. Contrairement à la carte de Patterson, elle contient donc une information de phase.

Nous avons utilisé le calcul d'une synthèse de Fourier pour faire apparaître l'ensemble des sites de fixations des lanthanides autour d'un modèle donné. Le coefficient utilisé pour le calcul de la synthèse est la différence de Bijvoet $D_{ANO} = |F(+)| - |F(-)|$. La phase utilisée pour calculer la carte est en retard de $\frac{\pi}{2}$.

La carte obtenue par synthèse de Fourier anomale peut également servir à déterminer une origine commune entre plusieurs jeux dérivés.

1.2.4 Un triste sire en ligne de mire, son seuil franchi, la triste Hélène de Troie devint folle : description des méthodes de phasage *de novo*

Au cours de ces quinze dernières années, les méthodes utilisant la diffusion anomale, c'est à dire la méthode SAD (Single-wavelength Anomalous Diffraction) et la méthode MAD (Multiple-wavelength Anomalous Diffraction), ont quasiment remplacé les méthodes qui utilisent exclusivement le remplacement isomorphe : les méthodes SIR (Single Isomorphous Replacement) et MIR (Multiple Isomorphous Replacement). Elles sont devenues les méthodes de choix pour la détermination *de novo* de la structure des macromolécules biologiques. 11530 structures sur les 64784 structures obtenues par cristallographie des rayons X recensée sur la Protein Data Bank répondent aux mots clés “anomalous”, “SAD”, “MAD”, “SIRAS” ou “MIRAS” (les méthodes “SIRAS” et “MIRAS” combinent l'utilisation de la diffusion anomale et le remplacement isomorphe). L'utilisation de la diffusion anomale a été portée par deux événements majeurs : le développement des sources synchrotron de seconde puis de troisième génération capable d'émettre un rayonnement X intense et de longueur d'onde ajustable d'une part et l'utilisation de protéines recombinantes séléniées comme méthode d'incorporation de diffuseurs anomaux d'autre part (Hendrickson *et al.*, 1990; Doublié, 1997).

1.2.4.1 Les méthodes de remplacement isomorphe, SIR et MIR

Les méthodes de phasage par remplacement isomorphe consistent à comparer les intensités des pics de diffraction d'un cristal natif et d'un cristal dérivé contenant des atomes lourds fixés dans le cristal.

Les atomes lourds insérés doivent avoir un numéro atomique nettement supérieur à ceux des atomes constitutifs de la protéine (S, N, C, O, H) afin que les intensités diffractées par le cristal dérivé diffèrent significativement de celles diffractées par le cristal natif.

Une condition essentielle pour la réussite de ces méthodes est l'isomorphisme des structures cristallines de la protéine native et de la protéine dérivée. L'insertion des atomes lourds sur la molécule ne doit ni modifier l'orientation des molécules de protéines dans le cristal ni modifier de manière importante les paramètres la maille cristalline. La densité électronique ρ_{PH} du cristal dérivé peut alors être considérée comme la somme de la densité électronique ρ_P de la macromolécule seule, identique à celle du cristal natif, et de la densité électronique ρ_H de la sous-structure des atomes lourds. Cette propriété se traduit sur les facteurs de structure par la transformée de Fourier et l'isomorphisme conduit à l'égalité :

$$F_{PH} = F_P + F_H$$

où F_P, F_{PH} et F_H sont respectivement les facteurs de structure du cristal natif, du cristal dérivé et de la sous-structure des atomes lourds dans le cristal dérivé.

Dans le cas de la méthode MIR avec deux cristaux dérivés isomorphes, les mesures des intensités diffractées par le cristal natif et les deux cristaux dérivés conduisent respectivement aux modules des facteurs de structure $|F_P|$, $|F_{PH1}|$ et $|F_{PH2}|$. Après avoir déterminé les coordonnées des atomes lourds à partir des cartes de différences de Patterson des deux cristaux dérivés, il est possible de calculer F_{H1} et F_{H2} en amplitude et en phase. Dans le plan complexe, F_P se trouve sur un cercle de rayon $|F_P|$ centré à l'origine alors que F_{PH1} et F_{PH2} se trouvent respectivement sur un cercle de rayon $|F_{PH1}|$ et d'origine $-F_{H1}$ et sur un cercle de rayon $|F_{PH2}|$ et d'origine $-F_{H2}$. L'intersection de ces trois cercles permet de déterminer la phase de F_P **[Figure 1.1]**.

La méthode SIR, qui utilise un cristal natif et un seul cristal dérivé, mène à deux solutions pour chaque phase résultante. Comme on a pu le voir, pour déterminer la phase des facteurs de structure sans ambiguïté, il faut au moins deux cristaux dérivés isomorphes différents. Dans ce cas (avec plus d'un cristal dérivé), on parle alors de "méthode MIR". C'est la méthode de phasage qui à été utilisée en 1954 pour résoudre la première structure de l'hémoglobine (Green *et al.*, 1954).

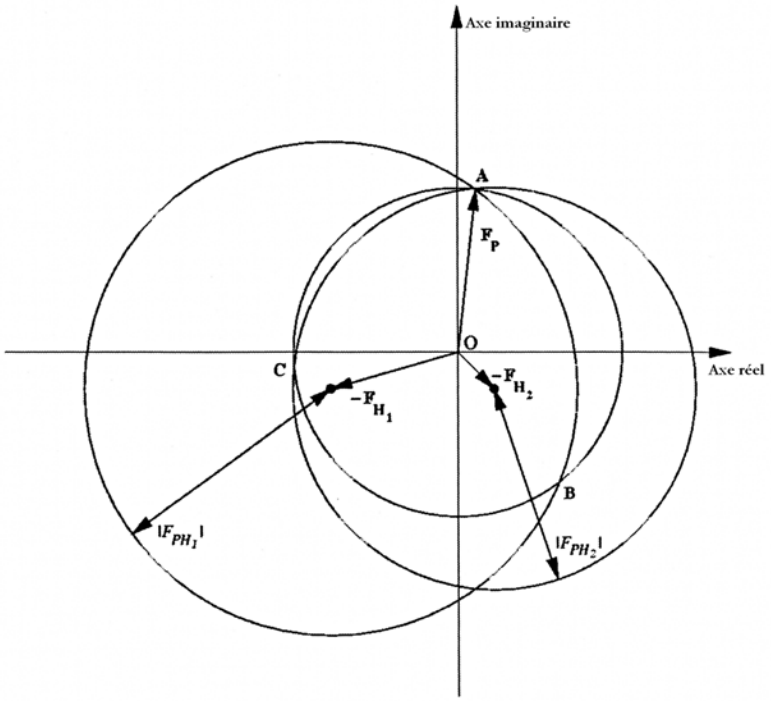


Figure 1.1 – Diagramme de Harker illustrant la détermination de la phase par la méthode MIR

1.2.4.2 Les méthodes exploitant la diffusion anomale, SAD et MAD

Par rapport aux méthodes de remplacement isomorphe, les méthodes SAD et MAD ont l'avantage de s'affranchir du problème de non-isomorphisme. En effet, ces méthodes utilisent les données de diffraction d'un cristal dérivé unique. Ces deux méthodes utilisent les différences anomales pour la détermination des phases. La première étape du phasage par ces méthodes est la localisation des sites de fixation des diffuseurs anomaux. Pour cela, il est nécessaire de calculer la carte de Patterson anomale du cristal dérivé (*cf.* ci-dessus). Les amplitudes $F(+)$ et $F(-)$ sont connues car déduites des intensités enregistrées lors de l'expérience de diffraction.

Si les coordonnées des diffuseurs anomaux sont déterminées, on peut non seulement calculer l'amplitude des contributions de ces atomes lourds au facteur de structure mais aussi leur phase. Connaissant les positions $\mathbf{x}_j$ des atomes lourds j dans la maille, on calcule F_H et F''_H, les contributions respectives des parties réelles et imaginaires de la contribution anomale au facteur de structure :

$$F_H = \sum_j ({}^0f_j + f'_j) \exp(i2\pi \mathbf{h} \cdot \mathbf{x}_j) \, et \, F''_H = i \sum_j f''_j \exp(i2\pi \mathbf{h} \cdot \mathbf{x}_j)$$

Comme la contribution des atomes diffusant anomalement peut s'écrire :

$$F_P = \sum_j f_j \exp(i2\pi \mathrm{h} \cdot \mathbf{x}_j) = |F_P| \exp(i\alpha_P)$$

où α_P est la phase. Etant donné la somme des contributions conduisant aux facteurs de structure :

$$F(+) = F_P + F_H + F''_H \, et \, \bar{F}(-) = F_P + F_H - F''_H$$

où $\bar{F}(-)$ est le complexe conjugué de $F(-)$. On en déduit :

$$F_P = F(+) - (F_H + F''_H) = \bar{F}(-) - (F_H - F''_H)$$

Le diagramme de Harker permet d'illustrer la détermination de la phase de F_P **[Figure 1.2]**.

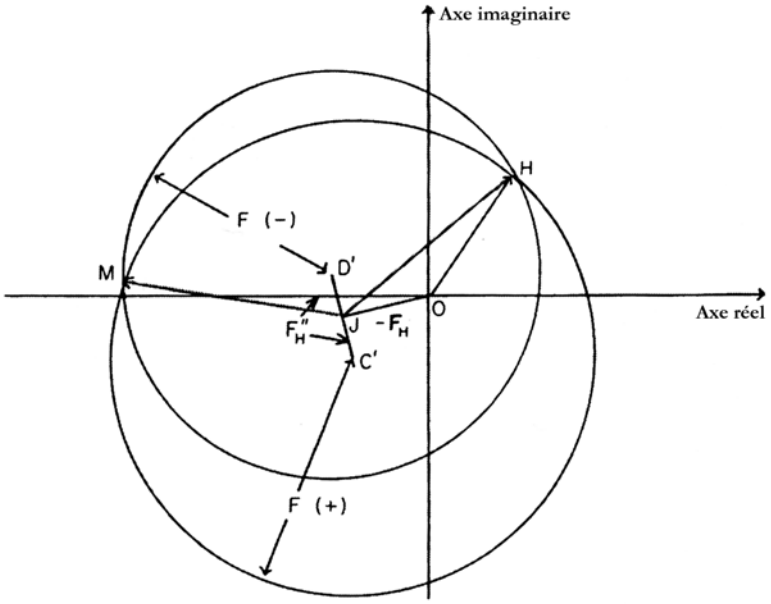


Figure 1.2 – Diagramme de Harker illustrant la détermination de la phase par la méthode SAD

Pour la méthode SAD, dans le plan complexe, les cercles de rayons $|F(+)|$ et $|F(-)|$ et d'origines respectives C', extrémité du vecteur $-(F_H + F''_H)$, et D', extrémité du vecteur $-(F_H - F''_H)$, se croisent en deux points d'intersection H et M. Les deux vecteurs résultants JH et JM correspondent aux deux solutions possible de la phase α_P de F_P.

Dans le cas de la méthode MAD, l'ambiguïté de phase est levée par une mesure réalisée à une autre longueur d'onde donnant d'autres valeurs de $|F(+)|$ et $|F(-)|$ conduisant au tracé de deux autres cercles se coupant à l'un des deux points H ou M.

De plus, avec la méthode MAD, une information supplémentaire peut être utilisée : la variation de la partie réelle du facteur de diffusion atomique f' avec la longueur d'onde. De la même manière que la différence anomale est définie à une longueur d'onde, on définit la différence dispersive entre deux longueurs d'ondes λ_1 et λ_2 : $D_{Dis} = \left|\tilde{F}^{\lambda_1}\right| - \left|\tilde{F}^{\lambda_2}\right|$ où $\left|\tilde{F}^{\lambda_i}\right|$ est le module moyen des réflexions $F(+)$ et $F(-)$ à la longueur d'onde λ_i. Les contributions au phasage des deux longueurs d'onde seront d'autant plus grandes que les différences dispersives seront élevées. Ce qui est le cas si les deux longueurs d'onde choisies conduisent à des valeurs très différentes de f'.

Historiquement, c'est en 1951 que Bijvoet *et al.* (1951) ont proposé que les variations de f' et f'' en fonction de la longueur d'onde pouvaient permettre de mimer un remplacement isomorphe. Ils ont ainsi ouvert la voie à l'utilisation de la diffusion anomale pour résoudre le problème des phases en cristallographie. Après qu'un formalisme pour la détermination des phases pour les cristaux dérivés non-centrosymétriques ait été développé (Okaya and Pepinsky, 1956) et que le signal anomal du fer aux

longueur d'ondes d'émission CoKα et NiKα fut utilisé pour déterminer les phases de plusieurs centaines de réflexions d'un cristal d'hémoglobine (Hoppe, 1975), c'est en 1981 que le pouvoir de phasage de ces méthodes fut démontré : la structure de la protéine crambine a pu être résolue avec le faible signal anomal du soufre à la longueur d'onde d'émission CuKα (Hendrickson, 1981). Avec l'avènement du rayonnement synchrotron, il est devenu possible de mesurer la résonance des raies blanches aux seuils d'absorption L des métaux de transition et des lanthanides (Lye *et al.*, 1980) et de travailler à la longueur d'onde correspondant au maximum du facteur f''. La méthode MAD est alors apparue comme une méthode de choix pour la détermination des phases (Phillips and Hodgson, 1980). Les bases algébriques de la détermination d'une solution unique par la méthode MAD furent posées indépendamment par Hendrickson (Hendrickson, 1979) et Karle (Karle, 1980). Ces bases algébriques ont ensuite été utilisées pour la résolution des premières structures par la méthode MAD (Kahn *et al.*, 1985).

1.2.4.3 Les méthodes combinées SIRAS et MIRAS

Les méthodes SIRAS (Single Isomorphous Replacement with Anomalous Scattering) et MIRAS (Multiple Isomorphous Replacement with Anomalous Scattering) tiennent compte, en plus des différences d'intensité des pics de diffractions observées entre un cristal dérivé et le cristal natif utilisé, des différences anomales des données de diffraction du cristal dérivé comme sources d'information de phase supplémentaire. Ces méthodes nécessitent donc l'enregistrement de données de diffraction issues d'un cristal natif et celles issues d'au moins un cristal dérivé contenant un atome lourd présentant une contribution anomale. D'une part, ces méthodes permettent donc de travailler en deux temps. D'autre part, si la résolution maximale imposée par les bords du détecteur à la longueur d'onde utilisée pour enregistrer les données de diffraction du cristal dérivé n'est pas assez élevée, l'enregistrement des données sur le cristal natif peut permettre d'atteindre la haute résolution. Le processus de phasage réussi, on utilise généralement ces méthodes pour obtenir uniquement la structure de la protéine native, le jeu de données anomal ne servant qu'au processus de phasage proprement dit. Toutefois, ces méthodes souffrent potentiellement du problème d'isomorphisme.

Pour ces méthodes, les atomes de lanthanide choisis initialement en tant que diffuseurs anomaux sont aussi intéressants. Ils fournissent en effet un fort signal isomorphe puisqu'ils possèdent entre 57 électrons (lanthane) et 70 électrons (lutétium).

1.2.4.4 Protocole retenu pour la détermination des phases par les méthodes SAD et MAD

Je décris ici le protocole et les programmes que j'ai utilisés pour la détermination des phases par les méthodes SAD et MAD.

Une fois les coordonnées des sites de diffuseurs anomaux déterminées, nous avons introduit cette information dans le programme *SHARP* (de La Fortelle, 1997) pour le calcul des phases expérimentales. Selon la méthode du maximum de vraisemblance, ce programme calcule la sous-structure que composent les atomes lourds dans la maille. Les coordonnées tridimensionnelles, le taux d'occupation et le facteur d'agitation thermique (isotrope ou anisotrope) des diffuseurs anomaux sont affinés ainsi que le facteur d'échelle, les écarts dus à un éventuel non-isomorphisme et les parties imaginaires et réelles du facteur de diffusion anomale. Après avoir déterminé un premier jeu de phases, le programme peut calculer des cartes résiduelles à partir desquelles d'éventuels sites anomaux supplémentaires peuvent être

détectés. L'information de phase est modélisée sous la forme des coefficients de Hendrickson et Lattman (Hendrickson and Lattman, 1970), coefficients qui décrivent la distribution de probabilités de phases.

Il est ensuite possible d'améliorer ces coefficients d'Hendrickson et Lattman par modification de densité via les programmes d'aplatissement de solvant *SOLOMON* (Abrahams and Leslie, 1996) et/ou *DM* (Cowtan and Main, 1996). En s'appuyant sur le pourcentage de solvant dans la maille cristalline calculé par le programme *TRUNCATE* de la *suite CCP4*, le programme SOLOMON détermine l'enveloppe de la protéine, limitant la zone de solvant, à partir des phases à basse résolution. Le nivellement de solvant consiste à fixer à une valeur constante positive la densité électronique correspondant au solvant. La correspondance d'histogramme consiste à ajuster la distribution des valeurs de la densité électronique observée à la distribution des valeurs qui est attendue pour la densité électronique à une résolution donnée.

Pour un phasage par la méthode SAD, je dirais que c'est à cette étape que réside le phasage expérimental. Comme nous l'avons vu, dans le cadre de la détermination des phases par la méthode SAD, la distribution de probabilités de phases est bimodale et apporte moins d'information qu'avec la méthode MAD qui mène à une distribution de probabilités de phases sans ambiguité. Ainsi, non seulement les modifications de densité permettent de lever l'ambiguïté en faisant ressortir la phase correspondant à la densité modifiée mais c'est aussi et seulement à l'issue de cette étape que l'expérimentateur peut savoir si la carte de densité électronique est interprétable ou non.

Enfin, via l'utilisation du programme de construction automatique *Arp/Warp* (Perrakis *et al.*, 2001), j'ai pu, non seulement juger la qualité de la carte de densité électronique expérimentale calculée à partir de ces phases grâce au critère "nombre d'acides aminés construits sur le nombre d'acides aminés total" mais aussi obtenir un modèle moléculaire de départ. Le programme a été utilisé avec ses paramètres par défaut, c'est à dire 10 cycles de construction automatique avec 10 cycles d'affinement pour cycles de construction. Le programme d'affinement alors appelé est celui de la *suite CCP4* : *Refmac*. Seule la séquence primaire à été utilisée pour la construction, je n'ai jamais utilisé de modèle ou d'une partie de modèle à étendre.

Au début de ma thèse, les premières reconstructions automatiques de modèles ont été réalisées avec le programme *Arp/Warp*. C'est donc dans un souci de reproductibilité et de comparaison que j'ai continué à utiliser ce programme tout au long de mes travaux. Mais malgré la détection d'un signal anomal prouvant la fixation des complexes, elle-même vérifiée par le calcul de la synthèse de Fourier utilisant les phases apportées par un précédent modèle, la qualité des cartes de densité électronique améliorées obtenues n'ont pas toujours permis au programme *Arp/Warp* de construire un premier modèle.

En cas d'échec de la construction automatique avec le programme *Arp/Warp*, j'ai utilisé un autre programme ayant le même rôle : le programme *Buccaneer* de la *suite CCP4* (Cowtan, 2006). La carte de densité électronique de départ améliorée par aplatissement de solvant et la séquence d'acides aminés alors utilisées ont été les mêmes que lors de l'utilisation du programme *Arp/Warp*. Cependant, pour le programme *Buccaneer*, seuls 5 cycles de construction automatique ont alors été utilisés, sans affinement sur les solutions intermédiaires. Le temps de calcul de ce programme étant déjà extrêmement rapide, l'avantage de l'utiliser selon cette procédure est que la solution peut s'obtenir en quelques minutes.

1.2.4.5 Evaluation de la qualité des phases expérimentales dans *SHARP*

Plusieurs indicateurs statistiques permettant d'évaluer la qualité des phases expérimentales sont calculés durant tout le processus de détermination et d'amélioration de phases. Dans cette partie, je donne la description de ceux que nous avons retenus pour illustrer la qualité des phases obtenues à l'aide des complexes de lanthanide.

"Phasing power" Le "phasing power" ("Ppower" ou "**Φp**") permet d'estimer l'efficacité du phasage en cours et permet d'anticiper l'issue de ce phasage. Il est mesuré pour les réflexions à phase non-contrainte ("$\boldsymbol{\varphi p}^{\neq isomorphes}_{sharp\,acen.}$", "$\boldsymbol{\varphi p}^{\neq anomales}_{sharp\,acen.}$", dans les tableaux) et à phase contrainte. Avec le diagramme de Harker en mémoire :

$\phi p^{\neq isomorphes} = \frac{|FH|}{|FPHi - |FP+FH||}$ avec $|FP+FH|$, vecteur somme des facteurs de structure des atomes lourds (FH) calculés, de la phase ($FHIH$) et des amplitudes moyennes (FP) de la protéine avec la phase. $|FPHi - |FP+FH||$ est l'erreur de fermeture calculée pour les différences isomorphes.

$\phi p^{\neq anomales} = \frac{|FH|}{||FPHi(+) - FPHi(-)| - 2.FHi.sin(\varphi)|}$ avec $|FPHi(+) - FPHi(-)|$, la différence anomale observée et $|2.FHi.sin(\varphi)|$, la différence anomale calculée où φ est la phase de la protéine. Calculée à partir des différences anomales, l'erreur de fermeture est :

$||FPHi(+) - FPHi(-)| - 2.FHi.sin(\varphi)|$.

Si les données et/ou l'isomorphisme ne sont pas parfaits ou encore si la position des atomes lourds n'est pas parfaitement définie, les différents cercles ne se croisent pas au même endroit : c'est l'erreur de fermeture. La solution n'est alors plus une unique valeur de phase mais une distribution de probabilités de phases $P(\varphi)$ que l'on modélise soit avec les coefficients d' Hendrickson-Lattman soit par le Facteur de Qualité "FOM".

"FOM" Nous évaluons la précision des phases expérimentales à l'aide du Facteur de Qualité "FOM" (Figure Of Merit), mesuré à chaque étape du phasage pour les réflexions à phase non-contrainte et à phase contrainte.

Pour un module de facteur de structure supposé connu :

$FOM = \frac{\left|\int_0^{2\pi} P(\phi) e^{i\phi} d\phi\right|}{\int_0^{2\pi} P(\phi) d\phi}$ où ϕ est la phase et $P(\phi)$, sa distribution de probabilité.

Ce Facteur de Qualité est le cosinus de l'erreur de phase. Les valeurs qu'il peut prendre sont donc comprises entre 0 et 1. Plus celle-ci est élevée, plus la distribution de probabilité des phases est réduite et plus cela traduit la garantie du succès du phasage. En effet, pour une phase connue et précise, on peut dire qu'une synthèse de Fourier ajoute une valeur de 1 pour une position d'un atome donné. Une valeur de $\mathbf{FOM}^{acen.}_{sharp} = 0{,}5$ correspond à une erreur de phase de 60°, erreur de phase que l'on considère comme acceptable. C'est pour cela que nous avons toujours prêté attention à la résolution pour laquelle la valeur du FOM calculé pour les réflexions à phase non-contrainte ("$\mathbf{FOM}^{acen.}_{sharp}$", dans les tableaux) était à peu près égale à 0,5. Cela nous a permis de savoir jusqu'à quelle résolution il était possible d'obtenir des phases suffisamment précises pour anticiper le succès, ou non, du phasage.

Dans le cas d'un phasage MAD, nous nous sommes aussi référés à la valeur du Facteur de Qualité calculée pour les réflexions à phase contrainte ("$\mathbf{FOM}^{cen.}_{sharp}$", dans les tableaux). Sa valeur reflète la contribution des atomes lourds vis à vis de la quantité totale d'atomes qui composent la protéine. Cela

permet d'avoir une idée du taux d'occupation du diffuseur anomal dans le cristal dérivé, donc d'avoir une estimation de la quantité d'atomes lourds fixés ou, d'une certaine manière, du taux d'occupation de ces atomes. Ici, les facteurs f'' et f' n'interviennent pas, seules sont comparées la quantité d'électrons qui composent la protéine avec celle des atomes lourds.

1.2.4.6 Evaluation de la qualité des phases expérimentales améliorées par *SOLOMON*

Les programmes d'aplatissement de solvant permettent d'améliorer les phases précédemment calculées. Des indicateurs statistiques permettent de suivre cette amélioration, toujours dans le but de fournir une estimation quant à l'issue du phasage : la construction du modèle.

"I(4)" Tout au long du calcul des poids attribués aux facteurs de structure modifiés, l'indicateur I(4), calculé par le programme *SIGMAA*, traduit la comparaison des phases des facteurs de structure des données expérimentales avec les phases des facteurs de structure de la carte aplatie.

De manière générale, le programme *SIGMAA* est utilisé pour combiner un jeu de phases calculées avec un jeu de phases déjà déterminées auparavant, phases traduites sous la forme de coefficients d'Hendrickson-Lattman (ref : Read, 1986). The program SIGMAA (Read, 1986) can be used to combine a set of calculated phases with a set of previously determined phases for which the phase probability profiles are held in the form of Hendrickson-Lattman coefficients.

It calculates weighted Fourier coefficients either from the calculated phase from a (partial) model structure, or by combining phase probabilities from isomorphous phases with those from one or more (partial) structures. Comprise entre 0 et 1, plus la valeur de l'indicateur I(4) est élevée à l'avant-dernier cycle d'aplatissement, plus cela signifie que la carte de densité électronique et le phasage ont pu être améliorés. *De part l'expérience, je suis en mesure d'indiquer qu'une valeur de I(4) supérieure à 0,60 à l'avant-dernier cycle de construction conduit à une carte de densité aplatie aussi bien interprétable à l'œil que par les programmes de construction automatique de modèle.*

Même si la valeur que prend cet indicateur à l'avant-dernier cycle est importante, l'évolution entre celle au premier cycle et celle à l'avant-dernier est aussi à prendre en considération. *Une valeur inférieure à 0,1 au premier cycle est synonyme d'une mauvaise carte de départ et donc de phases de départ imprécises.* De plus, dans le cas de la méthode SAD, cette première valeur que peut prendre le I(4) permet de discriminer la bonne main de la mauvaise, valeur bien plus élevée au premier cycle d'aplatissement s'il s'agit de la bonne main.

1.3 La diffusion anomale

Afin de résoudre ce "problème des phases", nous utilisons des méthodes de calcul de phases qui exploitent le fait que le facteur de diffusion atomique d'un atome donné varie fortement, en module et en phase, dans la région d'un seuil d'absorption de cet atome. C'est ce phénomène qu'on appelle "la diffusion anomale".

1.3.1 Présentation du phénomène

La théorie classique de la diffusion d'un rayonnement électromagnétique par un électron, appelée diffusion Thomson, repose sur le modèle de l'électron libre. Sous l'action du rayonnement incident, les

électrons des atomes émettent un rayonnement de même fréquence mais en opposition de phase (déphasé de 180°) par rapport au rayonnement incident. Cependant, dans un atome, les électrons des couches internes (K et L) sont fortement liés au noyau et ces électrons ne peuvent plus être considérés comme libres. Leur comportement sous rayonnement peut être décrit avec le modèle de l'oscillateur harmonique amorti, de fréquence de résonance ω_0 correspondant à la fréquence d'absorption de l'atome. Le facteur de diffusion atomique f pour un électron "lié" peut alors s'exprimer de la sorte :

$$f = \frac{\omega^2}{(\omega^2 - i\alpha\omega_0\omega - \omega_0^2)}$$

avec :

ω**,** la pulsation du rayonnement incident,

α**,** le facteur d'atténuation.

Pour un rayonnement incident d'une longueur d'onde voisine du seuil d'absorption considéré, la résonance implique une modification de la diffusion : c'est le phénomène de diffusion anomale. Alors que dans le modèle de l'électron libre, le facteur de diffusion atomique f est réel, dans le cas de la diffusion anomale, il devient un nombre complexe et s'exprime de la manière suivante :

$$f = {}^0f + {}^\lambda f' + i\,{}^\lambda f''$$

où :

${}^0f$ est la partie du facteur de diffusion atomique indépendante de la longueur d'onde,

${}^\lambda f'$ est la partie réelle de la contribution anomale au facteur de diffusion,

${}^\lambda f''$ est la partie imaginaire de la contribution anomale au facteur de diffusion.

Lorsqu'il y a absorption, le rayonnement incident est absorbé par les électrons des couches inférieures qui sont excités vers des états d'énergie plus élevée. Il en résulte un changement de phase entre le rayonnement incident et le rayonnement diffracté. La composante imaginaire du facteur de diffusion, f'', est reliée au coefficient linéaire d'absorption, $\mu(\lambda)$, par le théorème optique (Mie, 1908), qui peut s'écrire :

$$f''(\lambda) = \frac{M\mu(\lambda)}{2\lambda r_0 N_A \rho} = K\frac{\mu(\lambda)}{\lambda}$$

avec :

K**,** le coefficient de proportionnalité : $K = \frac{M}{2r_0N_A\rho}$,

$r_0 = \frac{e^2}{4\pi\varepsilon_0 m_0 c^2}$**,** le rayon classique de l'électron, ou longueur de diffusion Thompson,

λ**,** la longueur d'onde du rayonnement incident

M**,** la masse molaire de l'élément,

N_A**,** le nombre d'Avogadro,

ρ**,** la masse volumique de l'élément.

Cette relation permet d'obtenir les variations de f'' selon l'énergie du rayonnement à partir des variations du coefficient d'absorption.

Les variations de f'' avec l'énergie étant connues, on obtient f' à partir de l'équation de dispersion de Kramers-Kronig (Kronig, 1926; Kramers, 1927) :

$$f'(E) = \frac{2}{\pi} P \int_0^\infty \frac{E' f''(E') dE'}{E^2 - E'^2} dE'$$

avec :

P**,** valeur principale de Cauchy

Nous rappelons ici que, si nous utilisons maintenant le terme "énergie" pour qualifier le rayonnement incident plutôt que d'utiliser "sa longueur d'onde", c'est que ces deux composantes sont facilement calculables car liées par la relation du spectre électromagnétique :

$$E = \frac{hc}{\lambda}$$

avec :

E**,** l'énergie du rayonnement électromagnétique,

h**,** la constante de Planck ($h \simeq 6,626069.10^{-34} J.s$),

c**,** la vitesse de la lumière dans le vide ($c = 299792456\, m.s^{-1}$),

λ**,** la longueur d'onde du rayonnement électromagnétique.

Les valeurs calculées par la mécanique quantique pour f' et f'' ne sont valables qu'en dehors des seuils d'absorption. Au voisinage de ces seuils, elles doivent être déterminées de manière expérimentale (compte tenu de la variation de f'' en fonction de la longueur d'onde : $f''(\lambda) = K\frac{\mu(\lambda)}{\lambda}$).

Bien qu'approximatif, le modèle de l'oscillateur harmonique amorti **[Figure 1.3]** représente bien les variations des facteurs de diffusion au seuil d'absorption déterminées de manière expérimentale **[Figure 1.4]**.

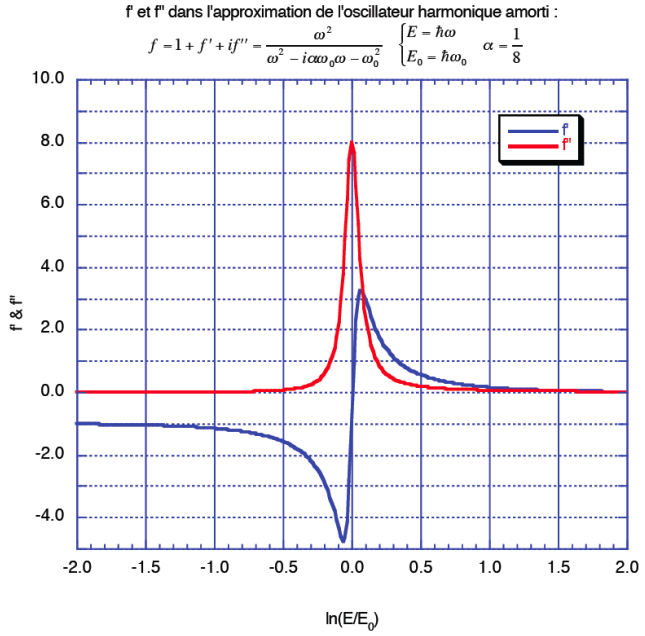


Figure 1.3 – Variation des parties réelles et imaginaires f' et f''du facteur de diffusion en fonction de l'énergie d'excitation, selon le modèle de l'oscillateur harmonique amorti. D'après Stelter (2005).

Les valeurs de f' et f'' sont donc déterminées à partir du spectre d'absorption de l'élément d'intérêt **[Figure 1.4]**, en effectuant par exemple une mesure de fluorescence. En effet, sous l'effet du rayonnement incident, les électrons excités quittent les couches internes pour les couches supérieures. Ceci crée des lacunes dans les couches internes qui sont alors de nouveau comblées par des électrons de couches de plus haute énergie. En venant combler ces lacunes, ces électrons émettent un rayonnement caractéristique de l'élément : la fluorescence. Tout en faisant varier la longueur d'onde autour du seuil d'absorption de l'élément, c'est cette fluorescence qui est enregistrée pour réaliser les spectres d'absorption. La fluorescence est proportionnelle à l'absorption. Les facteurs de diffusion anomaux f' et f'' sont calculés à partir de ces spectres d'absorption à l'aide du programme *CHOOCH* (Evans and Pettifer, 2001).

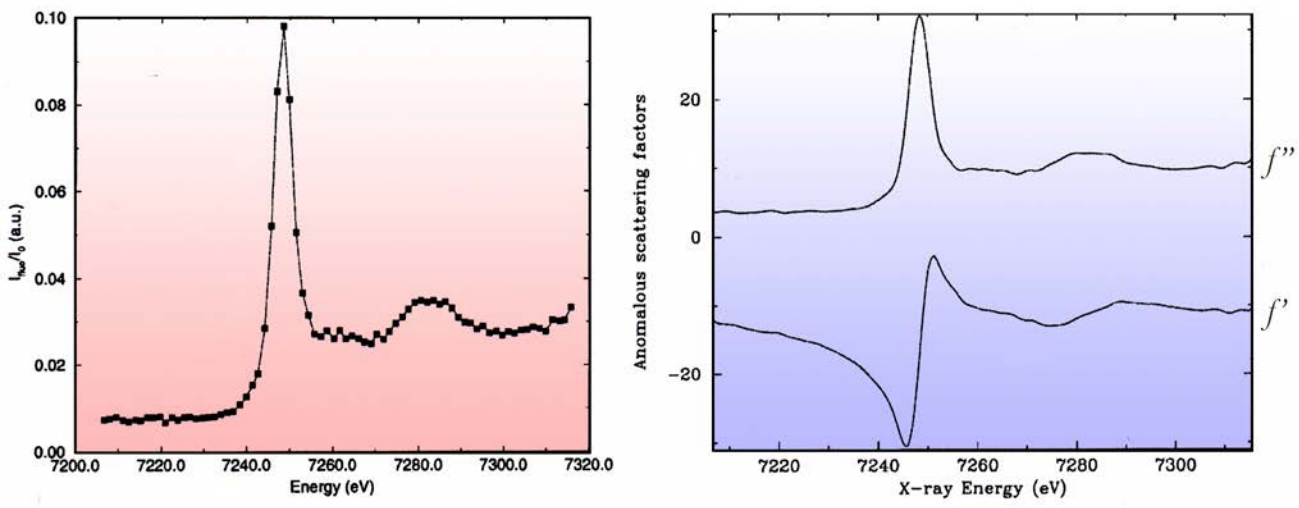


Figure 1.4 – Spectre d'absorption enregistré en mode fluorescence, sur la ligne BM30A de l'ESRF, sur une solution de 100 mM de complexe de gadolinium au seuil d'absorption L_{III} du Gd (à gauche). Facteurs de diffusion anomaux f' et f'' du gadolinium en fonction de l'énergie du rayonnement incident calculés par le programme *CHOOCH,* à partir du spectre d'absorption (à droite).

1.3.2 Les raies blanches

Le spectre d'absorption d'un atome présente parfois un pic marqué appelé "raie blanche". Ce phénomène se produit lorsque l'énergie du rayonnement incident correspond à l'énergie de transition de l'électron vers un état lié du continuum. La présence de raies blanches, pour des transitions depuis les couches K, L et M, est attribuée à l'existence d'un grand nombre d'états accessibles dans le continuum (Brown *et al.*, 1977). Comme le phénomène de raie blanche fait intervenir les orbitales moléculaires, l'allure des spectres d'absorption est fortement influencée par l'environnement moléculaire de l'atome considéré.

Dans le cas du seuil d'absorption L_{III} des atomes de lanthanide, ce sont les électrons de la couche L qui sont excités. Leur "raie blanche" est particulièrement marquée comme le montre le pic que présente leur spectre d'absorption, tel que celui de la **figure 1.4**. L'apparition des raies blanches aux seuils L semble liée à la présence d'états inoccupés dans la bande de valence *d*. Pour les atomes de lanthanide, la bande de valence *5d* étant totalement inoccupée, l'amplitude de la raie blanche aux seuils d'absorption L est identique pour tous ces atomes. L'amplitude et la position de la raie blanche peuvent être influencées par le degré d'oxydation de l'élément considéré. D'une manière générale, les ions lanthanides possèdent 9 atomes dans leur première sphère de coordination. L'énergie des orbitales moléculaires formées entre les lanthanides et les atomes coordonnés est donc particulièrement stable. Dans le cas de nos complexes de lanthanide, cette constance dans la géométrie de coordination des lanthanides peut donc expliquer pourquoi les raies blanches observées aux seuils d'absorption L_{III} ne varient que très peu selon le complexe étudié, ainsi le calcul des valeurs de f' et f'' peut se faire soit à partir d'une solution contenant l'atome considéré soit avec le cristal dérivé le contenant.

La plupart des lignes de lumière dédiées à la biocristallographie disposant d'un monochromateur accordable propose un accès à des longueurs d'onde de rayons X variant entre 0,7 Å et 2,0 Å. La **figure 1.5** montre l'évolution de la longueur d'onde des seuils d'absorption K, L et M en fonction du numéro atomique Z des éléments du tableau périodique. On voit que les seuils K des éléments légers situés entre

le manganèse Mn (Z = 25) et l'yttrium Y (Z = 39) sont facilement accessibles (λ = 1,89-0,73 Å). Les seuils d'absorption L des éléments situés entre le samarium Sm (Z = 62) et l'uranium U (Z = 92) sont également facilement accessibles (λ = 1,85 -0,72 Å).

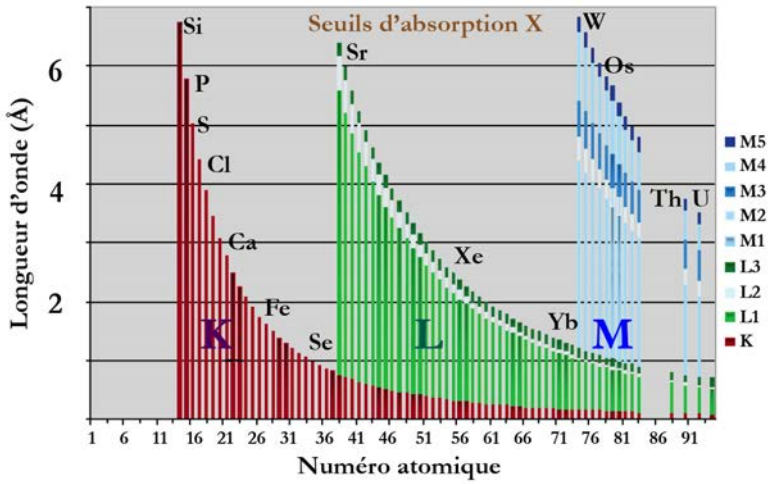


Figure 1.5 – Seuils d'absorption K, L et M pour les éléments stables du tableau périodique. D'après Chesne (2002).

Ces éléments sont donc potentiellement intéressants pour la détermination de structures par utilisation de la diffusion anomale. Cependant, tous ne présentent pas de raies blanches dans leur seuil d'absorption K ou L_{III}.

Pour les éléments ne présentant pas de raie blanche, la valeur de la partie imaginaire f'' du facteur de diffusion atomique est de l'ordre de 4 e^- pour la plupart des seuils d'absorption K et de 10 e^- pour les seuils d'absorption L (Joyce *et al.*, 2010). Ainsi, à leur d'absorption K, le sélenium (Se) et le manganèse (Mn) ont une contribution anomale f'' de l'ordre de 8 e^- et 10 e^-, respectivement.

A leur seuil d'absorption L_{III}, la contribution anomale f'' des atomes de lanthanide est d'environ 30 e^- **[Tableau 1.1]**.

Pour plusieurs raisons, on peut avoir recours à des expériences hors seuil, notamment si la longueur d'onde du seuil d'absorption de l'élément n'est pas accessible. On peut aussi vouloir enregistrer des données hors seuil si, dans le cadre d'une expérience MAD, on veut atteindre une résolution maximale. En effet, à résolution égale, les grandes longueurs d'onde imposent des angles de Bragg plus grands : la résolution maximale que l'on peut obtenir aux bords du détecteur, pour une distance cristal-détecteur donnée, est alors souvent inférieure à la limite de diffraction du cristal. Pour une position du détecteur identique, travailler à une longueur d'onde plus courte permet aussi de gagner en résolution. Cela est possible car, même hors seuil, on conserve une partie du signal anomal car f'' ne décroit que lentement pour $\lambda < \lambda_{seuil}$ **[Tableau 1.1]**.

Elément	**Eu**	**Gd**	**Tb**	**Yb**
Z	63	64	65	70
Seuil L_{III} (Å)	1,776	1,711	1,649	1,385
Seuil L_{III} (keV)	6,981	7,246	7,518	8,951
$f''_{L_{III}}$ (e^-)	$\simeq 30$			
$f'_{L_{III}}$ (e^-)	$\simeq$ -16			
$f''_{Cu_{K\alpha}}$ (e^-)	11,0	12,0	8,0	5,0
f''_{Se_K} (e^-)	6,3	6,7	7,1	9,3

Tableau 1.1 – Longueur d'onde des seuils d'absorption L_{III} des lanthanides utilisés lors de ces travaux de recherches et valeur approximative de la partie imaginaire *f"* de leur facteur de diffusion atomique à leur seuil d'absorption L_{III} respectif, au seuil d'émission Kα du cuivre (λ = 1,541 Å) et au seuil d'absorption K du sélénium (λ = 0,979 Å). Est aussi consignée la valeur approximative de la partie réelle f' du facteur de diffusion atomique de ces lanthanides à leur seuil d'absorption L_{III} respectif.

1.4 La brisure de la loi de Friedel

En l'absence de diffusion anomale, la loi de Friedel formule que les réflexions d'indices de Miller (hkl) et $(\bar{h}\bar{k}\bar{l})$ sont d'intensité égale. On considère que cette loi est respectée pour les molécules ne contenant que des atomes pour lesquels la contribution anomale est faible puisque, dans ce cas, on peut négliger la contribution anomale du facteur de diffusion atomique (cas d'un cristal natif). Dans ces conditions, le facteur de diffusion f_j de l'atome j et son complexe conjugué $\bar{f}_j$ sont égaux. Pour une paire de Friedel composée des réflexions $(hkl) \equiv \mathbf{h}$ et $(\bar{h}\bar{k}\bar{l}) \equiv \bar{\mathbf{h}}$, on a :

$$F(+) = F(\mathbf{h}) = \sum_j f_j \exp(i2\pi \mathrm{h} \cdot \mathbf{x}_j)$$

$$F(-) = F(\bar{\mathbf{h}}) = \sum_j f_j \exp(-i2\pi \mathbf{h} \cdot \mathbf{x}_j)$$

Si, pour tous ces atomes, $\bar{f}_j = f_j$, on voit que $\bar{F}(+) = F(-)$. Les deux intensités $I(+)$ et $I(-)$ d'une paire de Friedel sont en effet égales : $I(+) = F(+)\bar{F}(+) = F(-)\bar{F}(-) = I(-)$. Dans le cas d'un cristal natif ne contenant pas de diffuseurs anomaux remarquables, la loi de Friedel est donc respectée : les amplitudes d'une paire de Friedel $F_P(+)$ et $F_P(-)$ sont identiques.

Dans le cas d'un diffuseur anomal j, le facteur de de diffusion atomique f_j devient complexe : $f_j = {}^0 f_j + {}^\lambda f'_j + i^\lambda f''_j$.

Considérons ce diffuseur anomal introduit dans un cristal de protéine pour obtenir un cristal dérivé. Soit F_H et F''_H, respectivement les contributions des parties réelles et imaginaires du facteur de diffusion atomique au facteur de structure du diffuseur anomal. On a donc :

$$F_H = \sum_j ({}^0 f + f') \exp(i2\pi \mathbf{h} \cdot \mathbf{x}_j) \; et \; F''_H = i \sum_j f'' \exp(i2\pi \mathrm{h} \cdot \mathbf{x}_j)$$

Considérons les facteurs de structure du cristal dérivé, du cristal natif et des diffuseurs anomaux F_{PH}, F_P et $F_A = F_H + F''_H$, respectivement :

$$F_{PH} = F_P + F_A = F_P + F_H + F''_H$$

De part la contribution F_H et F''_H des diffuseurs anomaux, les amplitudes d'une paire de Friedel diffractées par le cristal dérivé deviennent différentes : $F_{PH}(+) \neq F_{PH}(-)$. La loi de Friedel est rompue **[Figure 1.6]**.

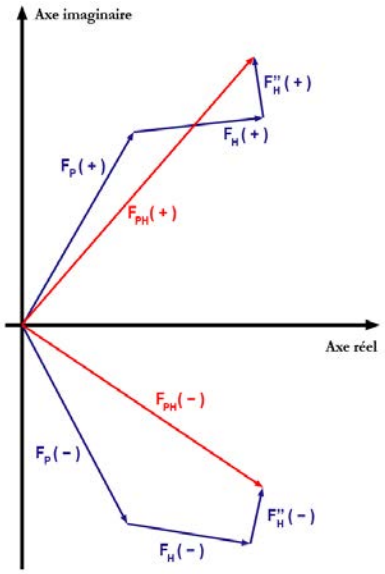


Figure 1.6 – Diagramme d'Argand illustrant la brisure de la loi de Friedel. F_{PH}, F_P, F_H et F''_H désignent respectivement les facteurs de structure : du cristal dérivé, du cristal natif, de la contribution réelle et de la contribution imaginaire des diffuseurs anomaux

On définit alors la différence anomale, ou différence de Bijvoet, par la différence des amplitudes d'une paire de Friedel : $D_{ANO} = |F(+)| - |F(-)|$. L'intensité des effets anomaux à une longueur d'onde donnée peut être estimée à l'aide du rapport de Bijvoet (Dauter, 2002; Girard *et al.*, 2003b) :

$$\frac{\langle |D_{ANO}| \rangle}{\langle |\bar{F}| \rangle} = \sqrt{\frac{2N_A}{N_P}} \frac{f''}{f_P} = \frac{1}{f_P} \sqrt{\frac{2\sum_j q_j^2 (f''_j)^2}{N_P}}$$

Où :

$\langle x \rangle \equiv \sqrt{\bar{x}^2}$ représente la racine carrée de la valeur quadratique moyenne de la grandeur x,

$|\bar{F}| = \frac{1}{2}(|F(+)| + |F(-)|) = \frac{1}{2}(|F(hkl)| + |F(\bar{h}\bar{k}\bar{l})|)$ est la valeur moyenne des amplitudes d'une paire de Friedel,

N_P est le nombre d'atomes de la protéine dans la maille possédant un facteur de diffusion moyen f_P,

N_A est le nombre de diffuseurs anomaux dans la maille dont la contribution imaginaire du facteur de diffusion atomique est f'' (ou f''_j pour un diffuseur anomal j ayant un taux d'occupation q_j).

Lorsque l'effet anomal est essentiellement dû à un seul type de diffuseur anomal, le rapport de Bijvoet est donc proportionnel à la partie imaginaire f'' du facteur de diffusion de ces atomes. Pour obtenir la différence anomale la plus importante, il faut se placer à la longueur d'onde où la valeur de f'' est maximale.

1.5 Evaluation du signal anomal et de la qualité des données de diffraction

La détermination des phases par les méthodes utilisant la diffusion anomale repose sur la différence d'intensité entre les réflexions d'une paire de Friedel. Cette différence étant souvent faible, il est indispensable d'enregistrer des données de diffraction suffisamment précises pour que le signal anomal soit significativement supérieur à l'imprécision des mesures. Lors de l'intégration des données de diffraction, première étape qui permet de mesurer toutes les intensités, plusieurs indicateurs statistiques rendent compte de la qualité des données de diffraction. Ils révèlent en outre l'importance de certains critères, décrits ci-après, dont il faut tenir compte lors de l'enregistrement des données.

1.5.1 L'importance du taux d'occupation du diffuseur anomal choisi

D'après la formule du rapport de Bijvoet, on voit que le signal anomal est non seulement dépendant de la valeur de f'' mais il est aussi dépendant du taux d'occupation q_j des sites de fixation des diffuseurs anomaux (Girard *et al.*, 2003b).

$$\frac{\langle|D_{ANO}|\rangle}{\langle|\bar{F}|\rangle} = \frac{1}{f_P}\sqrt{\frac{2\sum_j q_j^2 (f_j'')^2}{N_P}}$$

Il est donc important que les sites de fixation soient fortement occupés et que la longueur d'onde soit choisie de manière à obtenir la valeur de f'' la plus élevée possible si on veut déterminer les phases d'un grand nombre d'acides aminés (protéine de grande taille).

1.5.2 L'importance de la résolution des données

D'après la loi de Bragg, la résolution se définit en fonction de l'angle maximal de diffraction :

$$d_{réso} = \frac{\lambda}{2sin\theta_{max}}$$

En pratique, la résolution correspond à la plus petite distance entre les plans réticulaires du cristal pour laquelle des tâches de diffraction peuvent être enregistrées. Cependant, la formule décrite ci-dessus correspond à une définition dans l'espace réciproque et non dans l'espace direct. La résolution r, équivalente à $d_{réso}$ dans l'espace direct, qui correspond au plus petit détail visible dans la carte de densité électronique vaut $r = 0,715 d_{réso}$ (Blundell and Johnson, 1976).

Etant données les différentes contributions au facteur de diffusion atomique ($f =^0 f +^\lambda f' + i^\lambda f''$), seule le facteur de diffusion atomique de la protéine native ${}^0f$ décroît fortement avec l'angle de diffraction. Selon le rapport de Bijvoet que nous avons déjà défini, le signal anomal devrait alors être relativement plus élevé à plus haute résolution. Cet effet est cependant limité par l'agitation thermique de l'ensemble

des atomes et en particulier des diffuseurs anomaux, qui fait qu'à haute résolution les réflexions sont beaucoup moins intenses et que leur intensité est mesurée avec une précision moindre que celle des réflexions à basse résolution. On peut ainsi choisir de limiter la résolution d'un jeu de données de diffraction pour conserver un bon rapport signal anomal sur bruit.

1.5.3 L'importance du rapport signal sur bruit

Pour que le signal anomal soit significativement supérieur à l'imprécision des mesures, il faut améliorer l'exactitude des mesures, exactitude qui peut se mesurer par le rapport signal sur bruit :

$$\frac{I}{\sigma(I)}$$

Le rapport signal sur bruit dépend de la qualité du cristal mais il dépend aussi des conditions d'enregistrement : un temps de pose élevé et donc un nombre élevé de coups N_c par réflexion, améliore la précision statistique des données (le rapport signal sur bruit est approximativement proportionnel à $\sqrt{N_c}$).

Idéalement, le rapport signal sur bruit doit rester supérieur à 2 jusque dans la dernière tranche de résolution.

1.5.4 L'importance de la redondance des données

Elle correspond au nombre de fois qu'une réflexion unique a été enregistrée. Augmenter la redondance des données permet d'améliorer non seulement la précision statistique des mesures mais aussi et surtout le rapport "signal anomal sur bruit", en éliminant les erreurs systématiques dues aux artefacts de l'expérience (Dauter *et al.*, 2002). C'est pourquoi, pour un phasage utilisant le signal anomal, il est essentiel d'enregistrer des données de diffraction avec une redondance N élevée.

1.5.5 La complétude

Elle correspond au nombre de réflexions mesurées par rapport au nombre de réflexions théoriques :

$$C = \frac{N_1}{N_m} \; et \; N_1 = \frac{4}{3}\frac{\pi d^3}{V}$$

avec N_1 le nombre de réflexions théoriques, N_m le nombre de réflexions uniques mesurées, V le volume de la maille et d la distance entre les plans réticulaires.

La complétude est cruciale dans toutes les étapes qui font intervenir des synthèses de Fourier. Les réflexions manquantes affectant la qualité de la carte de densité électronique après l'étape d'affinement, il est nécessaire qu'elle soit la plus élevée possible. La valeur attendue de la complétude pour l'ensemble des données est comprise entre 95 et 100 %.

1.5.6 Les facteurs résiduels

Pour évaluer la qualité des données de diffraction enregistrées, les programmes d'intégration calculent plusieurs indicateurs de qualité.

La description de ces indicateurs de qualité est donnée ici avec $(hkl) \equiv \mathbf{h}$, les indices de Miller du réseau réciproque et N, la redondance associée à la réflexion d'indice $\mathbf{h}$.

1.5.6.1 Le facteur $\mathbf{R}_{merge}$

Indique l'accord entre les intensités équivalentes par symétrie et leur moyenne (Blundell and Johnson, 1976). Il se définit ainsi :

$$R_{merge} = \frac{\sum_{\mathbf{h}} \sum_{i} \left| I(\mathbf{h}) - \bar{I}_i(\mathbf{h}) \right|}{\sum_{\mathbf{h}} \sum_{i} \left| I_i(\mathbf{h}) \right|}$$

Avec :

$I_i(\mathbf{h})$**,** l'intensité de ma mesure i pour la réflexion d'indice $\mathbf{h}$,

$\overline{I}(\mathbf{h})$**,** intensité moyenne des réflexions équivalentes à cette réflexion $\mathbf{h}$.

Ainsi, ce facteur résiduel donne toujours des valeurs optimistes au début du traitement de données (lorsque la redondance n'est pas élevée) et augmente de manière asymptotique à mesure que le nombre de clichés de diffraction augmente.

1.5.6.2 Le facteur $\mathbf{R}_{pim}$

Cet indicateur, qui décrit aussi la précision des mesures, est plus informatif que le R_{merge} car il tient compte de la redondance des données N (Weiss, 2001).

$$R_{pim} = \frac{\sum_{\mathbf{h}} \left[\frac{1}{(N-1)} \right]^{\frac{1}{2}} \sum_{i}^{N} \left| I_i(\mathbf{h}) - \overline{I}(\mathbf{h}) \right|}{\sum_{\mathbf{h}} \sum_{i}^{N} \left| I_i(\mathbf{h}) \right|}$$

1.5.6.3 Le facteur $\mathbf{R}_{ano}$

Le facteur R_{ano} indique la précision des mesures si on considère la loi de Fridel brisée (jeu "anomal"). Il se décrit ainsi :

$$R_{ano} = \frac{\sum_{\mathbf{h}} \left| \overline{I^+}(\mathbf{h}) - \overline{I^-}(\mathbf{h}) \right|}{\sum_{\mathbf{h}} \left| \overline{I^+}(\mathbf{h}) + \overline{I^-}(\mathbf{h}) \right|}$$

avec $\overline{I^+}(\mathbf{h})$ et $\overline{I^-}(\mathbf{h})$, intensités moyennes d'une paire de Friedel.

1.5.7 Stratégie d'enregistrement des données de diffraction choisie au cours de ces travaux de thèse

L'ensemble des données de diffraction nécessaires à ce travail de recherche a été enregistré avec le rayonnement synchrotron sur les lignes de lumière MAD BM30A (*FIP : "French beamline for Investigation of Proteins"*), ID23-2 et ID29 du centre européen de rayonnement synchrotron *(ESRF :*

"European Synchrotron Radiation Facility", à Grenoble) et PROXIMA1, ligne de lumière du centre de rayonnement synchrotron SOLEIL (*"Source Optimisée de Lumière d'Énergie Intermédiaire du LURE"*, à Saint-Aubin).

Nous avons enregistré des données MAD à deux longueurs d'onde pour la majorité des cristaux dérivés. Nous avons toujours cherché à obtenir des données à haute résolution car la modélisation et l'affinement subséquent des complexes de lanthanide en dépend.

Les données MAD à deux longueurs d'onde ont donc été enregistrées de la manière suivante : enregistrement d'un jeu de données au seuil d'absorption L_{III} du lanthanide présent dans le cristal dérivé (en l'occurrence l'europium, le terbium, l'ytterbium ou le gadolinium) afin d'avoir un coefficient de diffusion anomale maximal ($f" \approx 30$ e^-).

Puis un second jeu de données est enregistré à la longueur d'onde correspondant au seuil K du sélénium (λ = 0,978 Å). Le choix d'enregistrer les données au seuil du sélénium peut s'expliquer par trois raisons.

- ▷ La première vient du fait qu'à cette longueur d'onde, nous avons accès à la haute résolution nécessaire pour l'affinement subséquent du complexe. Sur la ligne BM30A par exemple, le détecteur ADSC® Quantum 315r™ avancé jusqu'à la butée à 110 mm du cristal permet d'atteindre une résolution de d_{reso} = 0,98 Å. A cette distance minimale et à la longueur d'onde correspondant au seuil L_{III} de l'europium ou du gadolinium, la résolution imposée n'est plus que de d_{reso} = 1.93 Å.
- ▷ De plus, $f"$ ne décroissant que lentement pour $\lambda < \lambda_{seuil}$, à courte longueur d'onde, le coefficient de diffusion anomale reste à une valeur de $f" \approx 7$ e^- permettant de déceler un signal anomal suffisant. J'ai aussi pu utiliser ce signal anomal plus faible à cette longueur d'onde pour évaluer le "pouvoir phasant" des différents complexes en essayant de phaser, par la méthode SAD, avec ce seul jeu de données. Cela me permettait d'avoir une idée préliminaire du taux de fixation des complexes dans les cristaux dérivés.
- ▷ Enfin, il s'agit de la longueur d'onde pour laquelle la plupart des lignes de bio-cristallographie sont optimisées. Le faisceau est plus stable et brillant, ce qui me permet de diminuer le temps de pose par cliché de diffraction et donc de gagner du temps sur l'ensemble de l'enregistrement.

Sur BM30A, je n'ai jamais utilisé le positionnement automatique des appareils pour avoir accès aux seuils d'intérêt. Pour déterminer la longueur d'onde exacte du pic d'absorption, nous avons enregistré un spectre de fluorescence autour du seuil d'absorption en plaçant une boucle contenant la solution du complexe sur la tête goniométrique et nous avons utilisé le compteur MCA Roentec® et le logiciel de contrôle (la suite *XNemo*) installés sur la ligne. Les variations des facteurs $f"$ et f' ont été déduites du spectre de fluorescence à l'aide du programme *CHOOCH*. Cependant, lors des runs sur les lignes de lumière ID29 et ID23-2 de l'ESRF qui me sont moins familières bien que s'appuyant sur le même principe que celui de la ligne BM30A, j'ai utilisé le processus automatique de déplacement à la longueur d'onde du pic d'absorption choisi. Mais le processus accès aux seuils très opaque via cette automatisation a été source de pannes et d'incertitudes quant au positionnement exact au pic d'absorption.

Utilisation du robot de montage automatique de cristaux A l'heure où j'écris ces lignes, la plupart des lignes de lumière dédiées à la bio-cristallographie des seize synchrotrons mondiaux sont munies de chargeurs automatiques d'échantillons. Que ce soit le boîtier chargeur d'échantillons MK_{II} (collaboration entre l'EMBL, l'ESRF et la ligne BM14) ou le robot de montage automatique de cristaux

congelés et de plaque de cristallisation 96 puits CATS™ ("*Cryogenetic Automated Transfer System*", de la société française IRELEC®) installés sur la plupart des lignes de bio-cristallographie, ils facilitent et assurent le montage des échantillons ainsi que leur récupération.

Je prends à nouveau l'exemple de la ligne de lumière où j'ai passé le plus de temps : BM30A à l'ESRF. La présence du robot CATS™ sur la ligne de lumière change considérablement l'organisation générale d'un "run synchrotron" car, étant donné qu'il n'est plus nécessaire de rentrer dans la cabane d'expérience pour changer soi-même l'échantillon avec toute la procédure que cela implique, le gain de temps est appréciable. De cette manière, j'ai pu tester la qualité de diffraction de l'ensemble des cristaux congelés avant de ne retenir que les meilleurs pour les enregistrements complets. En outre, lorsque des acquisitions de données sur plusieurs cristaux dérivés contenant différents atomes de lanthanides sont prévues, impliquant donc l'accès à différents seuils d'absorption, le changement d'échantillons assuré et facilité par le robot permet la manipulation suivante. Plutôt que de garder un échantillon monté tout en changeant plusieurs fois la longueur d'onde, il est possible d'enregistrer d'abord les données avec tous les dérivés concernés par une même longueur d'onde avant de changer cette dernière. On minimise ainsi le nombre de changements de longueurs d'onde qui sont assez longs et quelque peu périlleux (ils peuvent parfois conduire à la perte du faisceau ou être sources de diverses pannes). N'utilisant pas le processus d'accès automatisé aux seuils de la ligne, procéder de la sorte m'a aussi permis de placer sur le goniomètre, alors disponible, un écran de fluorescence pour visualiser et améliorer la position et la brillance du faisceau en sortie de ligne après chaque changement de longueur d'onde. Le risque lié à cette approche est une éventuelle dégradation accrue des cristaux induite par le rayonnement.

Paramètres des enregistrements de données Pour les cristaux préalablement testés dont la diffraction n'était pas au delà de 1,90 Å sur la ligne BM30A et pour les données enregistrées sur les lignes de lumière ID29, ID23-1, ID23-2 de l'ESRF et Proxima1 du synchrotron SOLEIL, seules des données SAD au maximum de *f"* ont été enregistrées. En outre, pour les quatre dernières lignes de lumière précitées qui utilisent un élément d'insertion (ou "onduleur"), le faisceau a du être atténué afin d'éviter au maximum les dégâts d'irradiation et la saturation de la basse résolution.

Sur la ligne de lumière BM30A de l'ESRF, le temps d'enregistrement par image requis me permet de commencer à intégrer les intensités en cours d'enregistrement. Ceci m'a permis de changer d'échantillon lorsque le signal anomal alors évalué se révéla trop faible ou nul.

Tous les jeux de données de diffraction ont été enregistrés avec un pas angulaire $\Delta\phi$ de 0,5°, un temps de pose par image limitant toute saturation à basse résolution et avec une redondance élevée.

1.5.8 Protocole d'intégration des données de diffraction retenu

1.5.8.1 Obtention des intensités

Les intensités mesurées ont été intégrées avec le programme *XDS* (Kabsch, 1993). Une fois tous les paramètres initiaux implémentés, le profil des taches de diffraction est déduit des données expérimentales et ce avec d'autant plus de précision que le découpage angulaire de la rotation du cristal est fin. C'est pourquoi nous avons toujours travaillé avec un pas angulaire de 0,5°. Afin d'obtenir une intégration optimale des données, j'ai systématiquement appliqué cette méthode : affinement des paramètres cristallins et expérimentaux puis détermination du profil des taches de diffraction à partir du jeu de

données entier avant d'intégrer à nouveau les intensités avec pour paramètres de départ, ceux affinés au tour précédent. La table de bruit de fond est réalisée sur l'ensemble du jeu de données lors de la première itération. En conservant les résultats de chacune des itérations, il est aisé de suivre l'évolution du processus d'intégration, notamment via l'interface graphique "*xds_viewer*" de la suite logicielle ou à l'aide d'un script adéquat.

Nous précisons ici que nous avons constaté une amélioration considérable des résultats d'intégration entre l'utilisation de la version de décembre 2009 du logiciel et celle de décembre 2010 et ce, pour un même jeu de données et les mêmes paramètres initiaux utilisés. Certains jeux de données, traités avec la version de décembre 2009 (ou avec des versions ultérieures) qui n'avaient pu conduire à l'obtention de phases expérimentales ou qui avaient du être coupés en résolution afin d'obtenir des statistiques plus favorables ont finalement été exempts de tout défaut lors d'un nouveau traitement avec la version de décembre 2010 sans que cela soit lié à une augmentation du nombre de réflexions rejetées. Ceci peut s'expliquer par le soin apporté d'une part à l'analyse du bruit dû aux anneaux de glace et, d'autre part, à la réalisation de la table de bruit de fond. Dans la version de décembre 2010, ces deux routines donnent de bien meilleurs résultats lorsqu'on fait en sorte qu'elles soient réalisées sur l'ensemble des clichés de diffraction et pas seulement sur les premiers clichés.

La première étape du traitement des données cristallographiques est l'indexation de l'ensemble des tâches de diffraction pour déterminer les paramètres de maille et assigner à chaque réflexion un point du réseau réciproque. Ceci est effectué avec la routine *IDXREF* du programme *XDS*. Elle produit une liste de réseaux de Bravais possibles avec leur paramètres de maille associés. Le réseau de Bravais choisi pour l'étape suivante possèdera la plus haute symétrie et le meilleur accord de qualité.

La seconde étape consiste à déterminer l'intensité de chaque réflexion (routine *INTEGRATE*). Pratiquement, l'intensité de chaque pixel impliqués dans une tâche de diffraction sont sommées. Après une analyse de la forme des tâches, l'intensité est intégrée sur l'ensemble des pixels de la tâche.

La troisième et dernière étape est la mise à l'échelle par la routine *CORRECT*. Tout d'abord, les réflexions équivalentes, *i.e* celles mesurées plusieurs fois ou reliées par les opérations de symétrie, sont moyennées. Dans la théorie (cristal parfait), l'intensité est identique pour chacune de ces réflexions. En pratique, l'intensité de toutes ces réflexions équivalentes est moyennée et l'écart-type sur cette moyenne est noté R_{merge}. Cet écart-type est un des indicateurs de la qualité du jeu de données (voir ci-dessus).

Enfin, la totalité des intensités est mise à l'échelle d'une part en éliminant les signaux parasites et, d'autre part, en prenant en compte les fluctuations de l'intensité du faisceau incident de rayons X. Cette mise à l'échelle est réalisée par la routine *XSCALE*. A cette étape, on obtient une liste de réflexions uniques avec pour chacune d'elles la valeur moyenne de l'intensité (I) et sa déviation standard σ(I).

Autre indicateur statistique calculé par *XDS* que nous avons utilisé : SigAno : Evaluation des différences anomales estimées à partir des intensités d'une paire de Friedel $I(+)$ et $I(-)$ lorsque la loi de Friedel est brisée. Une valeur totale supérieure à 1,0 (calculée sur l'ensemble du jeu de données, pour toute la gamme de résolutions) indique une fixation du diffuseur anomal mais cela n'implique pas pour autant la réussite du phasage ultérieur.

1.5.8.2 Calcul des amplitudes des facteurs de structure à partir des intensités

A la fin de l'étape d'intégration réalisée avec le programme XDS, le fichier ASCII contenant les intensités de chaque réflexion n'est pas enregistré dans un format autorisant les diverses manipulations nécessaires pour la suite des événements. Nous avons pour habitude d'utiliser les programmes *COMBAT* et *SORTMTZ* pour le convertir en un tel format. Les différentes observations de chacune des réflexions sont mises à l'échelle avec le programme *SCALA* qui les convertit alors en intensités moyennes (Evans, 2005). Enfin, nous utilisons le programme *TRUNCATE* pour obtenir les amplitudes des facteurs de structure à partir de ces intensités mises à l'échelle. Ces quatre programmes font partie de la *suite CCP4 (Collaborative Computational Project, Number 4, 1994).*

De plus, les programmes *SCALA* et *TRUNCATE* calculent des indicateurs statistiques tels que les facteurs R_{merge}, R_{ano}, R_{pim}, le rapport signal sur bruit, la complétude, la redondance (ou "multiplicité") ainsi que la proportion théorique de solvant dans l'unité asymétrique. Indicateurs que j'utilise tout au long du manuscrit pour justifier la qualité des données cristallographiques et dont la description a été consignée dans cette section.

1.5.8.3 Estimation de la proportion de solvant dans l'unité asymétrique

En comparant le pourcentage théorique du volume de l'unité asymétrique occupé par les atomes que renvoie le programme TRUNCATE pour un nombre de molécules dans l'unité asymétrique imposé, on peut déterminer la part qu'occupent les atomes dans ce volume. Cette information sera nécessaire pour améliorer les phases ultérieurement. Par itérations successives, on cherche a atteindre une valeur de pourcentage de solvant logique pour une unité asymétrique de cristal biologique : entre 20 et 50 % de solvant.

1.6 Affinement de la structure et interprétations

L'étape de construction conduit à un modèle qu'il est nécessaire de compléter et d'affiner. Le protocole d'affinement repose d'une part sur l'évaluation de l'adéquation entre ce qui est construit et la carte de densité électronique calculée à partir des données expérimentales et, d'autre part, sur l'adéquation du modèle construit avec les propriétés physico-chimiques connues. Le but de tout cristallographe est d'obtenir le modèle le plus représentatif de la réalité.

L'évolution de la qualité du modèle est permise par le calcul de deux facteurs résiduels : Le R_{work} et le R_{free}.

1.6.1 Les facteurs résiduels utilisés lors du processus d'affinement

1.6.1.1 Le facteur R_{work}

$$R_{work} = \frac{\sum_{hkl} ||F_{obs}(hkl)| - |F_{calc}(hkl)||}{\sum_{hkl} |F_{obs}(hkl)|}$$

Ce facteur résiduel permet de suivre l'évolution de l'affinement en calculant l'adéquation entre le module des facteurs de structure observés $|F_{obs}(hkl)|$ avec le module des facteurs de structure calculé à partir du modèle construit $|F_{calc}(hkl)|$ et, pour toutes les réflexions.

1.6.1.2 Le facteur R_{free}

Il s'exprime de la même façon que le R_{work} . En revanche, il est calculé en utilisant une proportion de réflexions, choisies aléatoirement, qui ne sont pas incluses dans le processus d'affinement (typiquement 5 % des réflexions totales). Il constitue un indicateur non-biaisé de l'avancement de l'affinement.

1.6.2 Protocole d'affinement retenu pour nos travaux

Tous les affinements de structure ont été réalisés à l'aide du programme *Phenix.refine* de la *suite Phenix* (Adams *et al.*, 2010) dans ses versions v1.4 à v1.7.1 , en suivant le protocole décrit ci-après.

Toutes les valeurs statistiques et les cartes de densité électronique qui seront présentées dans ce manuscrit sont issues d'un affinement en mode "ml" simple. Partant d'un jeu anomal, nous avons cependant essayé de comparer ce type d'affinement à ce que pourrait apporter un affinement "mlhl", qui tient compte de l'information de phase experimentale (sous forme de coefficients d'Hendrickson-Lattmann). Mais aucune différence n'a été constatée (*cf.* Partie I, chapitre 2). Je consigne ici le protocole d'affinement qui a toujours été appliqué (sauf mention contraire) :

Le modèle de départ complété au maximum dans le programme *Coot* (Emsley *et al.*, 2010) à l'aide de la carte de densité électronique expérimentale améliorée par aplatissement de solvant par *Solomon* et/ou *DM*. Ce programme permet de visualiser les cartes de densité électronique et de reconstruire le modèle dans l'espace réel.

1. Avant de commencer l'affinement, trois possibilités s'offrent à nous :
   (a) Si le modèle initial provient du programme *Arp/Warp*, son utilisation par défaut conduit à la création d'un indicateur R_{free}. Ce dernier doit donc être transféré pour l'affinement avec *Phenix.refine* afin de s'assurer que les réflexions qui sont utilisées pour évaluer le R_{free} soient conservées.
   (b) Toujours dans le cas où le modèle initial provient du programme *Arp/Warp*, une fois complété, il est aussi possible d'effacer toute trace d'affinement de ce dernier avant de faire générer un nouveau R_{free} par *Phenix.refine* qui sera celui utilisé ensuite. Pour cela, il est nécessaire de retirer du modèle initial les molécules d'eau, les double-conformations et de remettre tous les facteurs d'agitation thermique aux valeurs de Wilson : 30,00 $Å^2$. C'est la méthode que j'ai le plus employée car elle m'a permis de m'affranchir de certains problèmes d'interprétation du R_{free} par *phenix.refine* que j'ai pu constater lorsque j'ai procédé ainsi).
   (c) Si le modèle initial a été construit par le programme *Buccaneer*, je l'ai utilisé sans modification supplémentaire en générant un R_{free} lors du premier tour d'affinement par *phenix.refine*.
2. Premier tour d'affinement : Calcul d'un R_{free} avec 5 % des réflexions exclues de l'affinement pour chaque tranche de résolution, si cas 1b) ou 1c).

(a) Recuit simulé. Permet de sortir du minimum d'énergie locale en augmentant virtuellement la température jusqu'à 5000 K typiquement, puis en la diminuant par pas de 100 K jusqu'à atteindre le seuil de 300 K,

(b) Affinement en corps rigide(s) des coordonnées. Cet affinement améliore la position dans la densité électronique de la molécule (protéine, monomère...) en la considérant comme un seul bloc, un tout rigide,

(c) Affinement des coordonnées atomiques pour chaque atome,

(d) Affinement de manière isotropique des facteurs d'agitation thermique, caractérisé par l'écart quadratique moyen. Soumis à une agitation, l'atome semble plus grand et la transformée de Fourier qui en résulte est plus large. On modélise ce phénomène avec le facteur d'agitation thermique.

3. Après avoir modélisé le complexe entier (atome de lanthanide et ligands) dans la carte de densité électronique $2F_o$-F_c et F_o-F_c, second tour d'affinement :

   (a) affinement des coordonnées atomiques,

   (b) affinement de manière isotrope des facteurs d'agitation thermique des atomes qui constituent la protéine et les ligands,

   (c) affinement de manière anisotrope des facteurs d'agitation thermique des atomes de lanthanide. La bibliothèque des contraintes géométrique correspondant au ligand du complexe a été obtenue avec le serveur *Dundee PRODRG2* (Schüttelkopf and van Aalten, 2004) ou le programme *phenix.elbow* de la *suite Phenix* (Zwart *et al.*, 2008) sans minimisation d'énergie intégrale.

4. Troisième tour d'affinement :

   (a) ajustement manuel du taux d'occupation du complexe entier,

   (b) affinement des mêmes paramètres que ceux décris à l'étape précédente.

   (c) Si d'autres ions ont pu être modélisés, l'affinement des facteurs d'agitation thermique de manière anisotrope peut être étendu à ceux-ci.

   (d) Construction automatisée des molécules d'eau.

5. Quatrième tour d'affinement :

   (a) ajustement manuel du taux d'occupation du complexe entier,

   (b) affinement des mêmes paramètres que ceux décrit à l'étape précédente,

   (c) construction manuelle des molécules d'eau,

   (d) construction des doubles conformations.

   (e) Si l'écart entre le R_{free} et le R_{work} a tendance à s'accroître (vers une valeur supérieure à 5 %), j'ai pu avoir recours à deux options supplémentaires :

      i. A cette étape, les facteurs d'agitation thermique étant déjà bien affinés, il est possible de faire une analyse TLS ("Translation, Libration, Screw") dont la description permettra de restreindre avec logique l'affinement ultérieur de ces facteurs. L'analyse TLS repose sur

la valeur du facteur d'agitation thermique des atomes qui composent la protéine, c'est pour cela qu'il est nécessaire qu'elle soit déjà affinée. Par la normalisation et la différenciation de la totalité des facteurs d'agitation thermique décrits, l'analyse TLS propose de partitionner une chaine principale en un certain nombre de segments qui sont alors considérés individuellement comme des corps rigides soumis aux mouvements vibrationnels suivants : la translation, la libration et l'hélicoïdalisation. Les atomes appartenant à un groupe TLS subissent la même variation. Selon la véracité avec laquelle ils expliquent le facteur d'agitation thermique observé (issue de l'affinement), une note est alors attribué à chacun de ces segments. Cette analyse met en évidence les mouvements locaux ou les mouvements des domaines protéiques, qui peuvent expliquer la flexibilité des macromolécules biologiques. Les analyses TLS ont été réalisées avec le serveur TLSMD (Painter and Merritt, 2006b,a).

ii. Lorsque la résolution était suffisamment haute ($\leq$ 1.30 Å), j'ai pris le parti d'affiner les facteurs d'agitation thermique de tous les atomes avec correction d'anisotropie, à la place du TLS.

iii. En dernier lieu et toujours dans le but de maintenir un écart raisonnable entre le R_{free} et le R_{work}, il peut être judicieux de laisser le programme déterminer le meilleur poids à accorder entre l'affinement des coordonnées atomiques et l'affinement des facteurs d'agitation thermique *(le calcul de ces poids requiert un temps de calcul plus long)*.

1.6.3 Les cartes de densité électronique

La carte $2F_o$ - F_c génère une densité électronique autour de l'ensemble du modèle (protéine, ions, ligands, molécules d'eau...). En complément, on utilise une carte F_o - F_c, nommée "carte différence", qui rend directement compte des désaccords entre le modèle et les données expérimentales.

1.6.4 Les facteurs de qualité d'affinement de structure

▷ Le diagramme de Ramachandran (Ramachandran and Sasisekharan, 1968). Ce diagramme permet d'évaluer la qualité de la géométrie de la chaine principale. Pour la majorité des acides aminés , seules les conformations énergiquement favorables sont permises pour les angles dièdres φ (axe N-C_α) et ψ (axe C_α-C) des liaisons peptidiques. Quatre zones distinctes sont définies en fonction des valeurs des angles φ et ψ : favorable, autorisée, généreusement autorisée et interdite. Ces zones ont été établies en tenant compte des paramètres stéréochimiques décrits par Morris *et al.* (1992) et à partir des paramètres propres aux structures connues (ces zones sont donc régulièrement mises à jour). Pour des structures autour de 2 Å de résolution, la géométrie générale du modèle est considérée satisfaisante lorsque 90 % des acides aminés se trouvent dans la zone autorisée. Ce diagramme peut être calculé par le programme *Procheck* (Laskowski *et al.*, 1993) ou *phenix.validate* de la *suite Phenix*.

▷ Les déviations standards. Certains paramètres physico-chimiques ont des valeurs dites "idéales" (longueur de liaisons, angles, planarité...) et le calcul de leurs déviations standards au sein du modèle est autant de critère de validité de la structure. Encore une fois, le programme *phenix.validate* de la *suite Phenix* permet une analyse détaillée du modèle pour tous ces paramètres.

Il convient aussi de prêter attention aux facteurs d'agitation thermique de l'ensemble des molécules affinées afin de juger l'homogénéité du modèle complet.

Les atomes de lanthanide ont été placés à l'aide de la carte de densité calculée à partir de la synthèse de Fourier anomale. La population des différents sites a été ajustée, par itérations successives, en analysant à chaque étape la carte différence F_o - F_c. Le taux de fixation du lanthanide et celui des ligands ont toujours été maintenus identiques et leurs facteurs d'agitation thermique ajustés par rapport à ceux des atomes voisins constituant la protéine. Le taux d'occupation des ligands du complexe est ajusté par rapport à celui du lanthanide qui sert de référence. Aussi, l'orientation des chaines latérales des acides aminés ont été vérifiées, tout comme les angles φ et ψ des liaisons peptidiques de ces derniers afin qu'ils puissent se trouver dans les régions favorables ou autorisées du diagramme de Ramachandran. Enfin, ayant disposé de données de diffraction à haute résolution, quelle que soit la structure affinée, j'ai toujours fais en sorte que l'écart entre le R_{free} et le R_{work} ne soit pas supérieur à 5 %.

1.7 Des cristaux dérivés présentant un signal anomal pour un phasage *de novo*

Différentes méthodes existent pour introduire des diffuseurs anomaux dans des cristaux destinés à un enregistrement de jeux de données de diffraction anomale. Il est aussi possible de se servir des diffuseurs anomaux naturellement présents dans les protéines.

1.7.1 Utilisation de protéines séléniées

La méthode de phasage *de novo* la plus utilisée avec le rayonnement synchrotron est la méthode MAD appliquée à des cristaux de protéines séléniées (Hendrickson *et al.*, 1990). La méthode consiste à exprimer la protéine d'intérêt en incorporant de la sélénométhionine ou, plus récemment, de la selenocystéine (Sanchez *et al.*, 2002; Strub *et al.*, 2003; Walden, 2010) au lieu des méthionines (Doublié, 1997). L'atome de soufre de cet acide aminé est donc remplacé par un atome de sélénium qui présente un contribution anomale f'' de 6 à 10 e^- dans son seuil d'absorption K, selon d'oxydation du sélénium. Les avantages de cette méthode sont l'occupation complète des sites de sélénium et l'accès au seuil facilité sur les lignes de lumière habituelles. L'inconvénient majeur de cette méthode est la mise en œuvre relativement laborieuse. Il faut exprimer la protéine en milieu sélénié, ce qui conduit souvent à de faibles taux d'expression. La protéine recombinante séléniée cristallise souvent dans des conditions modifiées par rapport à la protéine native. De plus, certaines protéines eucaryotes ne peuvent pas être exprimées dans des systèmes bactériens et certaines protéines ne contiennent pas d'acide aminé méthionine.

1.7.2 Utilisation de la diffusion anomale des atomes intrinsèques

Il est possible d'utiliser comme diffuseurs anomaux certains atomes naturellement présents dans les protéines. Les méthodes de phasage profitent alors de sites de fixation de diffuseurs anomaux fortement, sinon totalement, occupés et ne nécessitent pas de préparation de dérivé.

Plus de 20 % des protéines déposées dans la Protéin Data Bank appartiennent à la classe des métalloprotéines et contiennent naturellement des métaux comme le Cu, le Fe, le Mn, le Ni et le Zn. Les seuils d'absorption K d'un grand nombre de ces métaux sont accessibles avec le rayonnement synchrotron

et conduisent à des valeurs de f'' pour l'atome concerné de l'ordre de 4 e^-. Ces protéines se prêtent donc à des expériences MAD ou SAD.

il est aussi possible d'exploiter le signal anomal du soufre contenu dans les acides aminés méthionine et cystéine. Ainsi, la première structure de protéine a été résolue, en 1981, avec le signal anomal du soufre (Hendrickson, 1981). Depuis 1999, un certain nombre de structures ont été résolues par la méthode SAD en se servant uniquement du signal anomal du soufre (Ramagopal *et al.*, 2003; Lakomek *et al.*, 2009; Koch *et al.*, 2010; Doutch *et al.*, 2012). Toutes les protéines contenant des acides aminés cystéine ou méthionine contiennent du soufre. Ces deux acides aminés représentent en moyenne 4% des acides aminés. Il est possible utiliser la diffusion anomale du soufre à une longueur d'onde hors de son seuil d'absorption (le seuil K du soufre étant à 5,02 Å). Toutefois, la diffusion anomale du soufre à des longueurs d'onde habituelles est très faible. Comme f'' augmente avec la longueur d'onde utilisée, il faut trouver un compromis entre une valeur élevée de f'' et les problèmes liés à l'utilisation de rayons X mous. Une longueur d'onde optimale avec une ligne de lumière non-optimisée pour les grandes longueurs d'onde conduit à des rapports de Bijvoet très faibles. Utiliser ce faible signal anomal nécessite des données d'une très grande précision, les données doivent donc être enregistrées avec une redondance très élevée. Cette méthode ne s'applique donc, pour le moment, qu'à des cristaux qui diffractent très bien et qui résistent aux dégâts d'irradiation.

1.7.3 Utilisation de gaz nobles

Il est aussi possible d'obtenir des cristaux dérivés en introduisant des gaz nobles, xénon ou krypton, dans des cristaux natifs (Schiltz *et al.*, 2003). Ces cristaux sont dédiés aux méthodes de phasage SIRAS ou MIRAS. Pour cela, on soumet les cristaux à une atmosphère du gaz sous pression dans une cellule dédiée. Ces gaz étant relativement bien solubles dans l'eau, ils diffusent rapidement dans les canaux de solvant du cristal. Cette méthode permet ainsi d'obtenir des cristaux dérivés qui sont isomorphes aux cristaux natifs et qui peuvent présenter un nombre élevé de sites de fixation.

1.7.4 Introduction d'atomes lourds par trempage et co-cristallisation

Des cristaux dérivés de macromolécules biologiques peuvent être aussi obtenus en trempant un cristal natif dans une solution concentrée en diffuseur anomal ou en faisant croître les cristaux dans des gouttes de cristallisation contenant la protéine et le diffuseur anomal. Pour obtenir des cristaux dérivés par trempage ou co-cristallisation, on utilise généralement les sels de platine, d'or, de mercure, d'uranium ou de plomb (K_2PtCl_4, $KAu(CN)_2$, K_2HgI_4, $UO_2(C_2H_3O_2)_2$, $HgCl_2$, $K_2UO_2F_5$, *etc.*). Le seuil d'absorption L_{III} de chacun de ces éléments est accessible et permet d'avoir une valeur de f'' d'environ 10 e^-, sauf pour le platine et l'uranium qui présentent une raie blanche dans ce seuil avec une valeur de f'' supérieure à 20 e^-. Les temps de trempages peuvent être relativement longs (quelques heures ou plusieurs jours) et les concentrations en sels d'atomes lourds varient généralement entre 0,1 et 10 mM. Mais des trempages bien plus courts ("short cryo-soak") semblent donner des dérivés de meilleure qualité cristalline, avec un meilleur isomorphisme, qu'avec des trempage longs (Nagem *et al.*, 2001; Dauter, 2002). En outre, la connaissance de la séquence de la protéine peut aider à orienter le choix des sels d'atomes lourds à tester. Ainsi les sels de mercure se lient de préférence à des cystéines libres ou à l'azote des histidines et les sels de platine, aux cystéines, aux histidines ou aux méthionines (Garman and

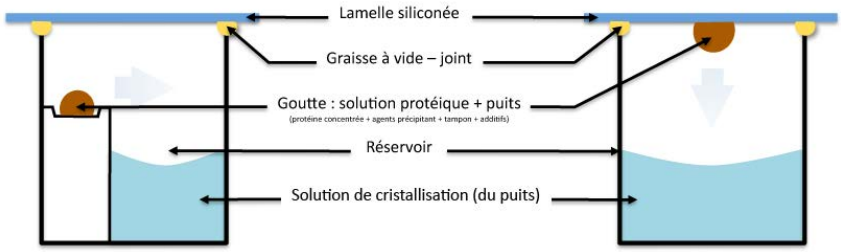
Murray, 2003). L'incorporation du diffuseur anomal dans la goutte de cristallisation pouvant introduire du non-isomorphisme, c'est avec le développement des méthodes SAD et MAD que la co-cristallisation trouva son essor.

Cependant, l'utilisation de ces sels d'atomes lourds n'est pas toujours aisée. Les sels sont généralement faiblement solubles et instables dans des solutions aqueuses, incompatibles avec de nombreuses conditions chimiques de cristallisation, souvent mal tolérés par les cristaux de protéine et hautement toxiques.

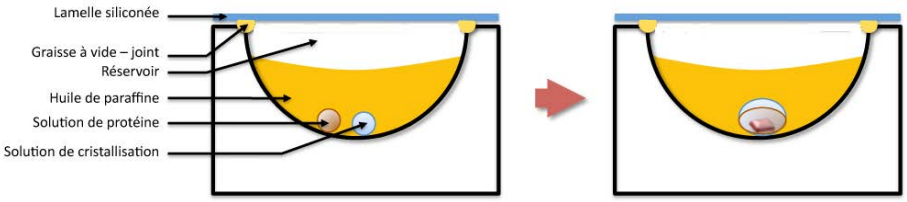
Depuis quatorze années, les diffuseurs anomaux qui font l'objet de nos travaux de recherche sont les atomes de lanthanides. Mais nous utilisons des complexes chimiques pour introduire facilement ces diffuseurs anomaux dans les cristaux de protéines par les méthodes de trempage et de co-cristallisation. Comme nous avons eu, en effet, recours à ces deux méthodes pour nos travaux, nous consignons ici le protocole général que nous avons utilisé pour obtenir des cristaux dérivés avec les complexes de lanthanide.

1.7.5 Protocole d'obtention de cristaux dérivés par trempage et par co-cristallisation avec les complexes de lanthanide

Obtention de cristaux natifs Toutes les protéines étudiées lors de cette thèse, sauf l'urate oxidase d'*Aspergillus flavus*, ont été cristallisées selon la méthode de la goutte suspendue **[Figure 1.7]**.


Lamelle siliconée
Graisse à vide – joint
Goutte : solution protéique + puits
(protéine concentrée + agents précipitant + tampon + additifs)
Réservoir
Solution de cristallisation (du puits)

(a) Schéma de la méthode de la goutte suspendue et de la goutte assise



(b) Schéma de la méthode "batch"

Figure 1.7 – Illustration des méthodes de la goutte assise, de la goutte suspendue et de la méthode dite "batch"

Le principe de la méthode est de mettre à l'équilibre, dans une enceinte hermétique, une goutte de quelques microlitres contenant la solution de cristallisation et la protéine face à un volume bien plus important (500 µL - 1 mL) de cette même solution mais sans la protéine **[Figure 1.7]**. La goutte est donc le mélange d'une solution concentrée en protéine et de la solution de cristallisation contenue dans le puits, elle-même étant un mélange d'agent(s) précipitant(s), de tampon et d'additifs éventuels. L'agent le plus volatil (l'eau) diffuse alors de la goutte vers le puits jusqu'à ce que leurs pressions partielles respectives

s'équilibrent, provoquant ainsi une diminution du volume de la goutte et donc une augmentation de la concentration de la protéine et des agents précipitants dans la goutte. Le diagramme de solubilité décrit l'évolution, dans la goutte, des concentrations de la protéine en solution et en agent précipitant **[Figure 1.8]**. On y distingue la zone de solubilisation, sous la courbe de solubilité et la zone de sursaturation. Cette dernière qui se subdivise en trois parties : la zone métastable, la zone de nucléation et le zone de précipitation. La goutte, dans son état initial, se trouve en un point sous la courbe de solubilité. Lorsque la diffusion commence, la concentration de protéine en solution [P] et en agent précipitant [AP] augmente dans la goutte pour atteindre la zone de nucléation. Dès lors, les cristaux se forment. Plus ils continuent de grossir, plus [P] diminue et la croissance s'arrête lorsque celle-ci est de nouveau au niveau de la courbe de solubilité.

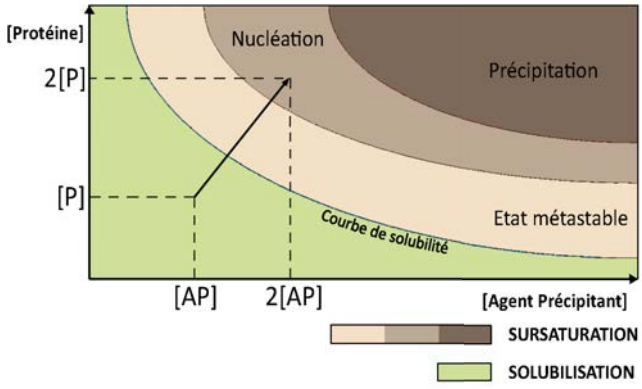


Figure 1.8 – Diagramme de phase

Les cristaux d'Urate oxidase d'*Aspergillus flavus* ont été cristallisés selon la méthode "batch" **[Figure 1.7]**.

Obtention des cristaux dérivés Il est possible de préparer des cristaux dits "dérivés" selon deux méthodes : le trempage et la co-cristallisation.

Le trempage, comme son nom l'indique, consiste à tremper un cristal natif dans une solution de composition identique à celle de la liqueur mère mais contenant aussi le complexe que l'on souhaite introduire dans le cristal. Les concentrations requises pour obtenir des cristaux dérivés avec une fixation suffisante du complexe sont variables car elles dépendent de la solubilité du complexe et des éventuelles dégradations constatées sur les cristaux natifs. Tous nos trempages dans la liqueur contenant les complexes de lanthanide ont été suivis d'un second trempage dans une même solution mais sans complexe (c'est le "back-soak", Garman and Murray, 2003) avant l'étape de cryo-protection.

La co-cristallisation consiste à faire croître un cristal de protéine en présence du complexe.

Nous avons essayé d'introduire les complexes à différentes concentrations par co-cristallisation avec toutes les protéines modèles disponibles et selon deux méthodes distinctes :

1. en ajoutant un volume de solution concentrée en complexe sur la goutte de cristallisation déjà constituée d'un volume de solution de protéine et d'un volume de solution de cristallisation. Dès l'ajout de ce volume, "à t_0", la protéine ainsi que les agents précipitants subissent le même facteur

de dilution. Leur concentration à t_0 est donc calculable, contrairement à celle atteinte une fois la goutte à l'équilibre, bien plus difficile à évaluer.

2. en dissolvant au préalable la protéine avec la solution de complexe, dans le cas des protéines lyophilisées (lysozyme et thaumatine) ou des protéines stables à haute concentration (eGFP, MalDH...). Cette méthode est avantageuse si on veut travailler avec les concentrations de complexe les plus élevées possibles, en particulier si la solubilité du complexe est faible.

Quelle que soit la méthode employée, il est important d'avoir une estimation du rapport des concentrations [complexe introduit] sur [protéine] pour chaque expérience afin d'estimer la concentration minimum de complexe à utiliser pour garantir sa fixation. En effet, si le complexe est en sous-stœchiométrie par rapport à la protéine, cela peut entrainer des taux d'occupation inférieurs pour les sites de fixation et faire échouer le phasage. Autant il est aisé de calculer ce rapport dans le cadre de la co-cristallisation, autant il est difficile d'avoir une estimation de ce rapport dans le cadre d'un trempage. On peut cependant toujours estimer la concentration de la protéine dans le cristal utilisé :

Soit un cristal de paramètre de maille a, b, c, α, β, γ et z, le nombre de molécules dans la maille.

$$C_{protéine} = \frac{n_{maille}}{V_{maille}} \; et \; n_{maille} = \frac{z}{N_A}$$

avec :

$V_{maille} = (a.b.c).\sqrt{1 + 2cos\alpha.cos\beta.cos\gamma - cos^2\alpha - cos^2\beta - cos^2\gamma}$**,** le volume de la maille qui doit être en dm^3 pour obtenir la concentration en $mol.L^{-1}$.

N_A**,** le nombre d'Avogadro : $6{,}022.10^{23}$ mol^{-1}

Ainsi :

$$C_{protéine} = \frac{z}{N_A.V_{maille}} = \frac{z}{N_A.(a.b.c).\sqrt{1 + 2cos\alpha.cos\beta.cos\gamma - cos^2\alpha - cos^2\beta - cos^2\gamma}}$$

Par exemple, dans le cas d'un cristal de lysozyme de blanc d'œuf de poule de groupe d'espace $P_4 2_1 2$, de paramètre de maille a = b = 78,0 Å, c = 38,0 Å, α = β = γ = 90,0° avec 8 molécules dans la maille, la concentration en protéine dans le cristal est de 57 mM. Pour un cristal natif de malate déshydrogénase de *Haloferax volcanii* de groupe d'espace C2, de paramètres de maille a = 144,370 Å, b = 82,028 Å, c = 112,32 Å, β = 101,626° avec 2 tétramères dans la maille, soit z = 8 protomères, $C_{protéine}$ = 10 mM. De la même façon que pour déterminer la concentration d'une protéine en solution, où seul le poids moléculaire et le coefficient d'extinction molaire du protomère sont pris en compte, nous prenons comme nombre de molécules dans la maille, le nombre de protomères. A z équivalent, les paramètres de maille conduisent à des concentrations très différentes.

Cependant, cette méthode n'est pas valable pour obtenir une estimation du rapport [complexe introduit] sur [protéine]. En effet, les volumes finaux correspondent respectivement au volume de la goutte de trempage et au volume du cristal introduit dans cette goutte de trempage. Afin de comparer la quantité de complexes de lanthanide introduits par trempage par rapport à la quantité de protéine, il faut raisonner en mole. Ainsi :

$$\frac{n_{complexe}}{n_{protéine}} = \frac{C^{dans\,la\,goutte}_{complexe}.V_{goutte}}{C^{dans\,le\,cristal}_{protéine}.V_{cristal}}$$

où,

$C_{complexe}^{dans\,la\,goutte}$**,** la concentration en complexes de lanthanide dans la goutte de trempage,

$C_{protéine}^{dans\,le\,cristal} = \frac{z}{N_A.V_{maille}} = \frac{z}{N_A.(a.b.c).\sqrt{1+2cos\alpha.cos\beta.cos\gamma-cos^2\alpha-cos^2\beta-cos^2\gamma}}$**,**

V_{goutte}**,** le volume de la goutte de trempage,

$V_{cristal}$**,** le volume du cristal, en dm^3 utilisé pour le trempage, déterminé à partir des dimensions du cristal mesurées au préalable.

Par exemple, compte tenu des dilutions, si la concentration en complexe dans une goutte de trempage de 2 µL $= 2.10^{-6}$ dm^3 est de $2{,}7.10^{-3}$ $mol.L^{-1}$, la quantité de complexe dans la goutte est de $5{,}4.10^{-9}$ moles. Si le cristal utilisé pour le trempage est un cristal de lysozyme de blanc d'œuf de poule avec des dimensions de 200x200x200 µm, la concentration en protéine dans le cristal est donc de 57 mM et le volume du cristal est de 8.10^{-9} dm^3. La quantité de protéine introduite dans la goutte de trempage est donc de $4{,}56.10^{-10}$ moles. Le rapport des quantités de complexe sur quantité de protéine indique qu'il y a donc 11,8 fois plus de complexes que de protéines pour l'obtention d'un cristal dérivé. Pour un même volume de goutte de trempage, de concentration en complexe dans la goutte et un cristal natif de malate déshydrogénase de *Haloferax volcanii*de même dimension, ce rapport prend une valeur de 66. On peut ainsi savoir si la volume de la goutte de trempage est suffisant par rapport au volume du cristal pour que le complexe ne soit pas en sous-stœchiométrie par rapport à la protéine dans le cristal.

Cette méthode de calcul nécessite donc de prêter attention d'une part aux dimensions des cristaux utilisés pour le trempage et d'autre part, au volume de la goutte de trempage.

Cryoprotection des cristaux Lors de leur exposition aux rayons X durs, les cristaux subissent des dégâts d'irradiation qui les dégradent au cours du temps. A l'origine, les dégâts d'irradiation sont dus à l'absorption des rayons X par la matière, arrachant des électrons aux atomes : c'est l'ionisation primaire. Certaines liaisons chimiques, comme les ponts disulfure, se retrouvent brisées. Aussi, des radicaux libres peuvent se former et diffuser dans le cristal. A température ambiante, les radicaux ainsi formés sont à l'origine de réactions radicalaires en chaîne qui dégradent certains acides aminés (Garman, 2010). Le fait de travailler à température cryogénique réduit la diffusion des radicaux et limite ainsi les dégâts associés.

Pour congeler les cristaux, il convient de choisir un cryoprotectant qui ne dégrade pas les cristaux et qui assure une vitrification suffisante. On choisi le cryoprotectant en fonction de la liqueur mère dans laquelle sont apparus les cristaux. Par exemple, pour des cristaux apparus dans une liqueur mère contenant du PEG, le PEG 400, l'éthylène glycol ou le glycérol sont des cryoprotectants de choix. Dans le cas de cristaux obtenus en présence de sulfate d'ammonium, on peut aussi choisir comme cryoprotectant le sucrose ou le saccharose. Si le cryoprotectant n'est pas adapté, il y a formation de glace cristalline lors de la congélation. Celle-ci crée non-seulement des contraintes mécaniques qui abiment le cristal mais elle altère aussi les clichés de diffraction, y apparaissant sous forme de diagramme de poudre. Le cryoprotectant entraîne, lui, la formation de glace vitreuse, amorphe, qui diffuse les rayons X sans conduire à un diagramme de diffraction de poudre.

Tous les cristaux étudiés au cours de nos travaux ont été transférés dans leur solution cryoprotectrice respective dont la composition correspond, sauf mention contraire, à celle de la liqueur mère à laquelle a été ajouté un cryoprotectant à la concentration requise. Les transferts sont réalisés à l'aide d'une boucle

et les temps de trempage sont de l'ordre de la seconde. Le cristal est ensuite congelé dans l'azote liquide à 100 K.

1.8 Utilisation des lanthanides en tant que diffuseurs anomaux

1.8.1 Les lanthanides, des diffuseurs anomaux sous-utilisés lorsqu'il s'agit de la préparation de cristaux dérivés

Comme nous l'avons précédemment indiqué, le sélénium est le diffuseur anomal le plus utilisé. Comparons le pouvoir de phasage des atomes de lanthanide à celui des atomes de sélénium a travers le rapport de Bijvoet décrit en 1.4. Si l'on suppose des taux d'occupation identiques pour les diffuseurs anomaux, et si l'on considère que les données de diffraction sont enregistrées au seuil d'absorption correspondant (la contribution imaginaire f'' vaut environ 10 e^- pour le sélénium et 30 e^- pour les lanthanides), le rapport de Bijvoet montre qu'un site lanthanide totalement occupé permettra de déterminer la structure d'une protéine de masse moléculaire neuf fois plus importante, comparé à un site sélénium totalement occupé.

L'utilisation des ions lanthanides en tant que diffuseurs anomaux a été proposée depuis que leur fixation dans le site de fixation calcium de certaines protéines a été démontrée. Dans ce cas précis, l'atome de lanthanide se substitue à l'atome de calcium (Horrocks and Sudnick, 1981). Cette particularité a été exploitée lors des premières expériences MAD (Kahn *et al.*, 1985). Comme nous l'avons indiqué, les ions lanthanide possèdent une coordination particulière pouvant accueillir de 3 à 11 atomes dans leur première sphère de coordination, avec une coordination privilégiée de 9. Ainsi, hormis dans le cas des protéines à calcium, la coordination complète est difficile à satisfaire à la surface d'une protéine : seuls les acides aminés aspartate et glutamate sont favorables à ce type de coordination avec leur fonction acide (mais un acide aminé ne peut, à lui seul, compléter la sphère de coordination du lanthanide). Aussi, l'utilisation de sels de lanthanide endommage généralement les cristaux de protéines.

Il existe diverses solutions pour introduire des lanthanides dans les cristaux de protéines afin de limiter leur détérioration. Il a été proposé d'utiliser des sels de lanthanides via la méthode de trempage rapide ("short cryo-soak") (Dauter *et al.*, 2000; Nagem *et al.*, 2001). Cette méthode consiste à tremper pendant un temps très court, inférieur à la minute, des cristaux dans une solution très concentrée en sel de lanthanide (au delà du molaire). Mais cette méthode provoque aussi une dégradation des cristaux. Suite à la publication initiale sur l'utilisation des complexes de lanthanide "IRM" (Girard *et al.*, 2003b), complexes qui seront décrits dans le chapitre suivant, la création d'une liaison covalente entre des complexes de lanthanide dérivés de précédents et la protéine, par l'intermédiaire d'une fonction réactive thiol, a été proposée (Purdy *et al.*, 2002). La fonction thiol permet la création de ponts disulfure avec les cystéines libres de la protéine. Plus récemment, la structure de l'ubiquitine a été résolue par phasage expérimental grâce à la greffe d'une étiquette à lanthanides sur la protéine (Silvaggi *et al.*, 2007). Cette étiquette ("lanthanoid binding tag") peut fixer un ou deux atomes de lanthanide. Elle est directement exprimée à la surface de la protéine puisque sa séquence d'ADN est introduite dans le gène qui code la protéine par les techniques de biologie moléculaire classique (Allen and Imperiali, 2010; Barthelmes *et al.*, 2010). Mais cette méthode est assez lourde à mettre en œuvre et l'encombrement stérique et la flexibilité qu'introduit l'imposante étiquette peuvent rendre difficile la croissance cristalline.

Afin d'introduire des lanthanides dans les cristaux de protéines, l'approche de l'équipe est d'utiliser des complexes de lanthanide très solubles qui se fixent aux protéines via des liaisons non-covalentes (*cf.* chapitre suivant).

1.8.2 La luminescence des lanthanides

Comme nous allons évoquer à de nombreuses reprises la luminescence de certains complexes de lanthanide, j'introduis ici brièvement la luminescence intrinsèque des atomes de lanthanides.

Les cations trivalents des lanthanides ont la configuration électronique (Xe) $4f^n$, avec n variant de 1 (pour le Ce^{3+}) à 14 (Lu^{3+}). Les transitions des électrons f sont responsables des propriétés optiques des ions de lanthanides. La luminescence des lanthanides est caractérisée par une longue durée de vie des états excités, de l'ordre de la milliseconde. C'est l'interdiction des transitions f-f qui est responsable de la longue durée de vie des états excités mais aussi de la faible photoexcitation directe des ions (Binnemans and Görller-Walrand, 1995). Pour des ions lanthanide en solution aqueuse, le rendement de luminescence est très faible. Les molécules d'eau de la première sphère de coordination donnent lieu à la désexcitation par voie non-radiative des états excités par couplage vibronique à des états vibrationnels des liaisons O-H et N-H. La liaison des molécules d'eau entraîne également un fort raccourcissement de la durée de vie de la luminescence.

De tous les ions lanthanide, seuls les ions Eu^{3+}, Tb^{3+}, Gd^{3+}, Ce^{3+}, Dy^{3+} et Sm^{3+} sont luminescents de manière significative en solution aqueuse et à température ambiante (Horrocks, 1993). Seuls l'europium et le terbium présentent, respectivement, des raies d'absorption et d'émission dans le domaine de l'ultraviolet et du visible.

La faiblesse des coefficients d'absorption des lanthanides est généralement contournée en coordonnant l'ion avec un ligand qui contient un chromophore. Le chromophore, qui sert d'antenne, absorbe de la lumière et transfère de l'énergie sur l'ion qui luminesce. Le ligand sert également à exclure des molécules de solvant de la sphère de coordination de l'ion, ce qui empêche la désexcitation non-radiative de l'ion. En augmentant ainsi l'absorption de lumière et le rendement de luminescence, on obtient un signal fortement amplifié qui permet de détecter facilement le complexe lanthanide-antenne luminescent.

Chapitre 2

Utilisation des complexes de lanthanide en bio-cristallographie

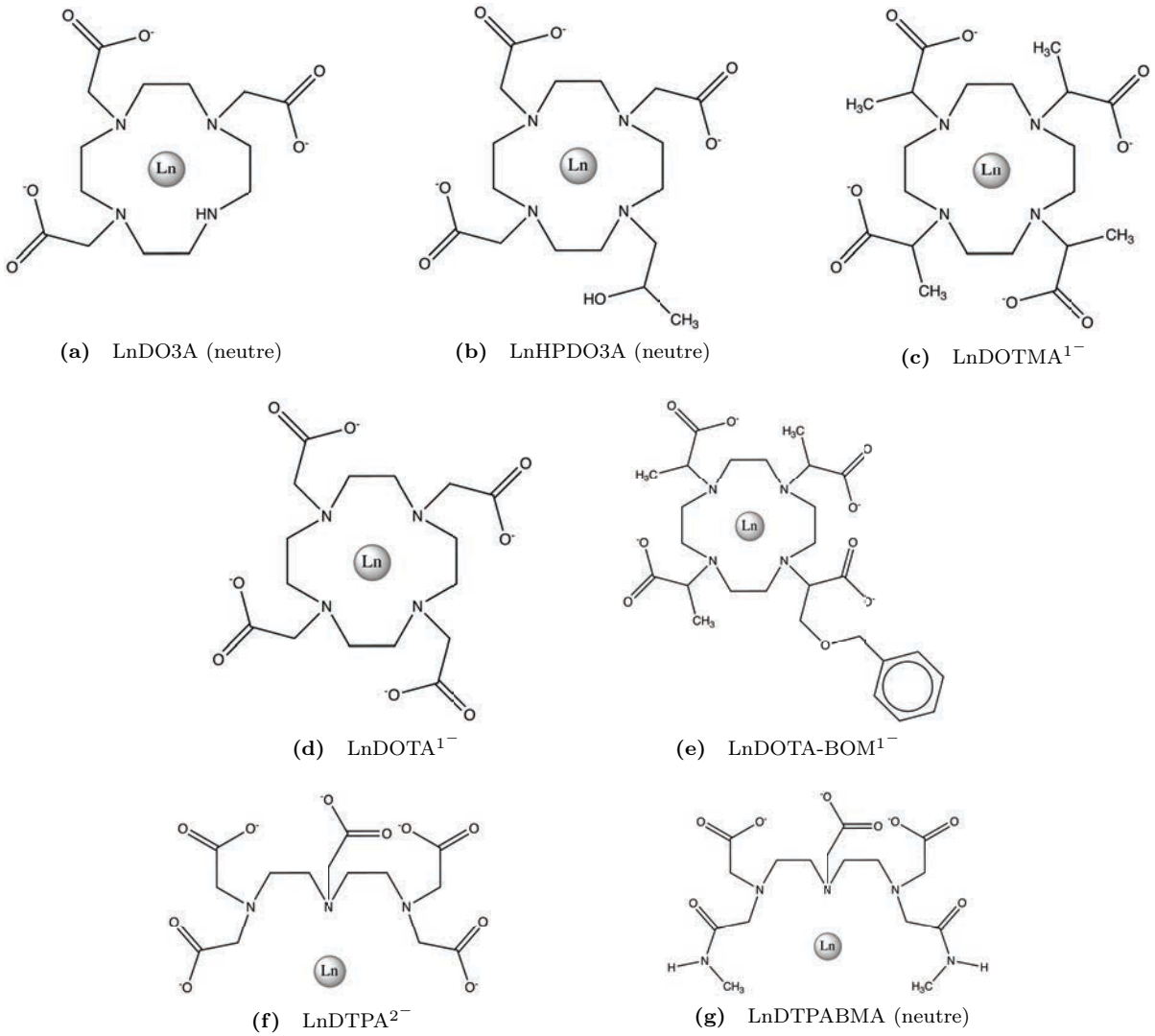
Afin d'introduire les lanthanides dans les cristaux de protéine, l'approche de l'équipe est, depuis quatorze années, d'utiliser des complexes de lanthanide. Ces complexes sont tous constitués d'un ligand qui chélate un ion lanthanide. Très solubles, ils peuvent être introduits dans les cristaux par trempage ou par co-cristallisation. L'ion lanthanide n'étant pas libre, les complexes de lanthanide ne sont pas nocifs et peuvent être utilisés sans précaution particulière. Dans la majorité des cas, c'est le ligand qui assure la fixation du complexe à la protéine. Nous récapitulerons dans ce chapitre l'ensemble de nos observations sur la fixation de ces complexes de lanthanide. Puis je rappellerai que certains complexes permettent d'exploiter les propriétés de luminescence des lanthanides en jouant le rôle d'antenne, propriété que nous utilisons pour détecter la fixation de ces complexes dans les cristaux avant l'étape de diffraction.

2.1 Les complexes de lanthanide issus de l'imagerie médicale

Historiquement, les complexes développés pour l'imagerie par résonance magnétique (que nous surnommerons "complexes IRM") ont été les premiers à être utilisés par l'équipe.

2.1.1 Présentation des complexes IRM

Les quatre complexes commerciaux GdHPDO3A, GdDOTA, GdDTPA et GdDTPABMA ainsi que des complexes dérivés de ces derniers ont été testés pour obtenir des cristaux dérivés **[Figure 2.1]**.



(a) LnDO3A (neutre) **(b)** LnHPDO3A (neutre) **(c)** $LnDOTMA^{1-}$

(d) $LnDOTA^{1-}$ **(e)** $LnDOTA\text{-}BOM^{1-}$

(f) $LnDTPA^{2-}$ **(g)** LnDTPABMA (neutre)

Figure 2.1 – Complexes issus de l'imagerie par résonance magnétique

Définition des ligands

DO3A acide 1,4,7,10-tetraazacyclododecane-1,4,7-triacétique, masse moléculaire théorique de 343,36 $g.mol^{-1}$,

DOTA acide 1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacétique, 400,39 $g.mol^{-1}$,

HPDO3A acide 10-(2-hydroxypropyl)-1,4,7,10-tetraazacyclododecane 1,4,7-triacétique, 401,44 $g.mol^{-1}$,

DOTMA acide 1,4,7,10-tetramethyl-1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacétique, 456,49 $g.mol^{-1}$,

DOTA-BOM acide (phenylmethoxy)methyl-1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacétique, 562,61 $g.mol^{-1}$,

DTPA acide ethylene-tri-amine pentaacétique, 388,31 $g.mol^{-1}$,

DTPABMA acide 1,3-bis-methylamide ethylene-tri-amine pentaacétique, 416,41 $g.mol^{-1}$.

Les ligands sont tous des cages qui peuvent chélater un ion Ln^{3+}. A part le ligand linéaire DTPA et ses dérivés, tous sont dérivés du macrocycle tétra-azacyclododécane ("cyclen").

Ces complexes sont très solubles dans l'eau (> 500 mM) et stables sur une très large gamme de pH (de pH 2,0 à 11,0). Comme le ligand assure l'essentiel de la coordination du lanthanide (6 à 8 coordinations), l'affinité élevée du ligand pour le lanthanide conduit à des complexes extrêmement stables dans des solutions fortement concentrées en sels ou en ions divalents. Ces complexes sont donc bien adaptés aux conditions physico-chimiques que l'on rencontre au cours d'expériences de cristallisation de protéines.

Comme l'indique **la figure 2.1**, certains de ces complexes ont une charge globale nulle (électriquement "neutre"), d'autres sont chargés négativement. Les valeurs de pKa des complexes étant situées autour de 2 (Bianchi *et al.*, 2000) aucune protonation notable n'a lieu dans les conditions physico-chimiques de cristallisation habituelles. Ces complexes sont introduits soit par co-cristallisation, soit par trempage et s'utilisent à forte concentration ($\approx$ 100 mM) sans que cela perturbe la cristallisation ou détériore les cristaux (Girard *et al.*, 2003b).

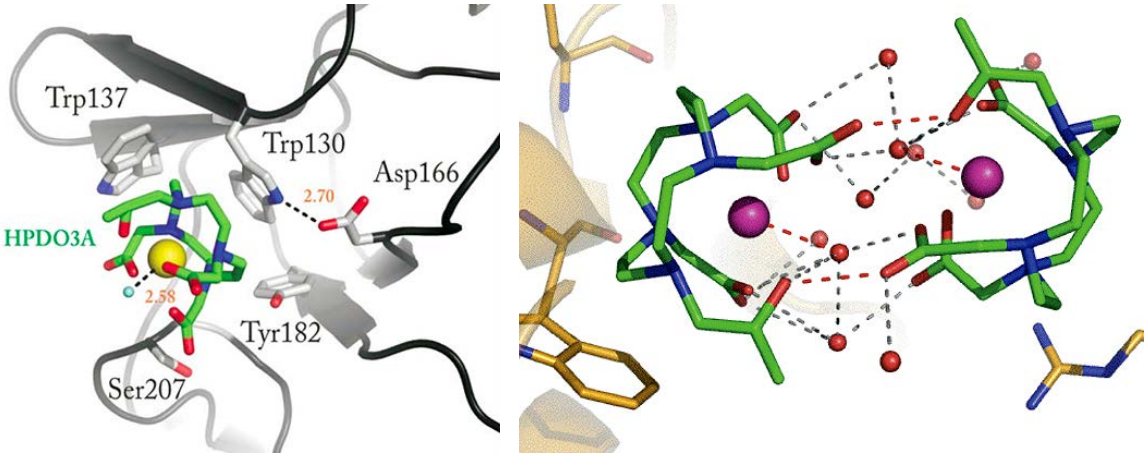
2.1.2 Etat de nos connaissances sur le mode de fixation des complexes de lanthanide IRM

Dans ce paragraphe, je ne ferai pas une liste exhaustive des différents modes de fixation de l'ensemble des complexes IRM en notre possession mais je me focaliserai sur ceux qui ont été le plus couramment utilisés.

2.1.2.1 Le complexe LnHPDO3A, pour des interactions hydrophobes

Le LnHPDO3A est une molécule électriquement neutre qui laisse entrer une molécule d'eau dans la première sphère de coordination du lanthanide. Le ligand HPDO3A présente une face hydrophobe et une face hydrophile. Elles constituent respectivement le macrocycle et les fonctions carboxyles avec la molécule d'eau. Ce complexe fait partie de ceux pour lesquels le mode de fixation est le mieux compris car, lorsqu'il se fixe avec un taux d'occupation supérieur à 50 %, le même mode de fixation a toujours été observé.

Ce mode de fixation implique un contact hydrophobe entre le macrocycle du ligand et des acides aminés hydrophobes de la protéine **[Figure 2.2]**. Cette interaction de van der Waals est renforcée par un réseau de liaisons hydrogènes impliquant les fonctions carboxyles, des molécules d'eau et des acides aminés au voisinage du site de fixation.



(a) Mode de fixation du HPDO3A observé dans les cristaux de protéine CbpF, **?**

(b) Mode fixation du HPDO3A observé dans les cristaux de lysozyme de blanc d'œuf de poule, d'après Stelter (2005).

Figure 2.2 – Mode de fixation du complexe LnHPDO3A

A la surface des protéines, nous avons toujours observé que, lorsque la fixation est forte, deux molécules de complexe s'associent, les deux molécules étant en contact par leur face hydrophile qui

participent au réseau de liaisons hydrogènes **[Figure 2.2]**. Dans ce cas, la distance entre les les deux atomes de lanthanide est d'environ 6 Å.

En "couvrant" les zones hydrophobes, le complexe LnHPDO3A forme un écran au solvant. L'énergie totale du système physico-chimique protéine+LnHPDO3A est donc plus faible. Le caractère énergiquement défavorable de présenter des zones hydrophobes à la surface de la protéine est d'autant plus important que la force ionique du solvant est importante. Ceci est en accord avec la forte fixation du LnHPDO3A lorsque les conditions de cristallisation contiennent des fortes concentrations en sels.

2.1.2.2 Le complexe LnDO3A, pour les interactions ioniques

Le complexe LnDO3A est aussi une molécule dont la charge globale est nulle. Sa structure est similaire au LnHPDO3A : ce complexe possède le même macrocycle, donc la même face hydrophobe, et une face hydrophile. La différence avec le LnHPDO3A réside dans le groupement hydroxy-propyl qui est absent chez le complexe LnDO3A. Cette absence conduit à la présence de deux molécules d'eau dans la première sphère de coordination. Bien que la structure entre les deux complexes soit très similaire, le complexe LnDO3A présente un tout autre mode de fixation.

Le LnDO3A se fixe sur les acides aminés aspartate et glutamate par une liaison ionique entre la fonction acide de ces acides aminés et l'ion lanthanide **[Figure 2.3]**. En chassant une des deux molécules d'eau, la fonction carboxylate de ces acides aminés fournit une coordination plus stable au lanthanide, qui, dès lors, devient similaire à ce que l'on observe avec les autres complexes neutres.

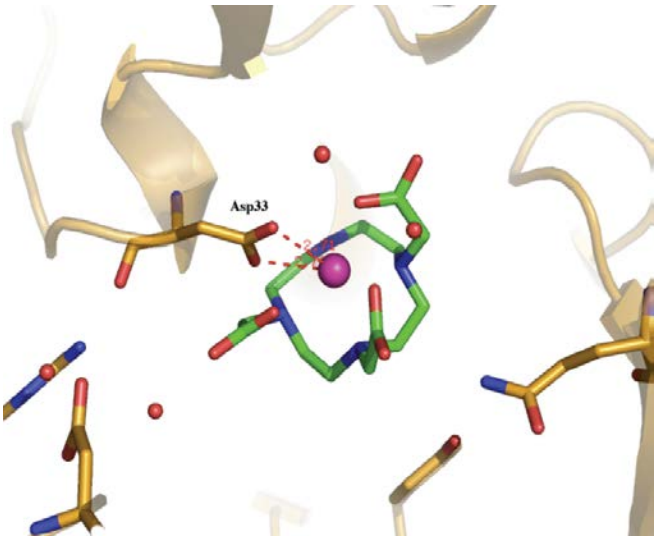


Figure 2.3 – Mode de fixation du complexe LnDO3A à la surface de la protéine YGGV (collaboration avec C. Abergel, IGS, Marseille), d'après Stelter (2005)

Du point de vue gain énergétique, le complexe LnDO3A diffère du complexe LnHPDO3A. En effet, le gain énergétique est en priorité en faveur du complexe qui, par le remplacement d'une des molécules d'eau, fournit une coordination plus stable au lanthanide sans pour autant déstabiliser la protéine puisque les acides aminés aspartate et glutamate sont généralement impliqués dans des ponts salins et des liaisons hydrogènes. Etant donné que c'est l'ion qui assure l'interaction principale avec la protéine et non plus le ligand, on retrouve les statistiques de fixation observées avec un sel de lanthanide : des

sites de fixation multiples dont le taux de fixation est de l'ordre de 50 % et une fixation plus ou moins indépendante des conditions de cristallisation (mais suffisante pour un bon phasage).

2.1.2.3 Les autres complexes IRM

Bien que les grandes lignes des modes de fixation soient comprises pour les complexes LnHPDO3A et LnDO3A, il reste parfois des cas de protéines où aucune fixation n'est observée, ce qui indique des modulations dont les causes ne sont pas encore bien comprises. Ce constat est encore plus vrai pour les autres complexes IRM testés. Bien qu'ils aient permis de résoudre un certain nombre de structures, leur mode de fixation varie d'une protéine à l'autre et il a toujours été souvent délicat d'expliquer la fixation ou l'absence de fixation de ces complexes.

Au cours de ses travaux de recherche, Meike Stelter a réalisé une étude systématique de 7 complexes IRM sur 11 protéines différentes. Les structures de ces protéines ont pu être résolues par phasage expérimental avec ces complexes. Dans ce paragraphe, je dresse le bilan des observations de Meike Stelter.

Le LnDO3A a permis de phaser 7 structures sur 8 protéines testées, le LnHPDO3A, 5 sur 9 testées, le LnDTPABMA, 4 sur 9 protéines testées et le LnDOTA-BOM 4 sur 5 protéines testées. Le complexe LnDO3A s'est donc montré le plus "universel".

L'étude du mode de fixation de ces complexes a révélé qu'aucune relation directe n'apparaît entre les conditions physico-chimiques (pH, précipitant, point isoélectrique de la protéine) et une éventuelle fixation privilégiée d'un complexe particulier. De manière générale, la fixation des complexes engendre peu de changement au niveau de la structure protéique, des paramètres de maille et de la forme cristalline. Pour certains dérivés la liaison du complexe engendre des mouvements de chaîne latérale de résidus en interaction directe avec le complexe. Elle résulte généralement de multiples interactions faibles, impliquant le plus souvent un réseau de molécules d'eau coordonnées entre le complexe et la surface protéique et elle semble dépendre de la forme de la surface protéique. Les complexes se lient souvent dans des cavités plus ou moins prononcées, formées par la surface de la protéine (thaumatine) ou par l'empilement cristallin (lysozyme), ce qui indique l'importance de facteurs stériques pour la fixation. Les complexes chargés négativement ont plus tendance à se lier sur des parties de surface chargées positivement que les complexes neutres. Aussi, les complexes semblent se lier de préférence selon une orientation qui accorde leur forme à la forme de la surface.

La fixation des 7 complexes testés peut combiner différents types d'interaction :

1) L'interaction directe entre l'ion Ln^{3+} et des atomes chargés négativement des résidus de la surface protéique (atome d'oxygène de la chaîne latérale des résidus Asn, Gln, Asp ou Glu).

2) Des ponts salins ou des liaisons hydrogènes directes entre les atomes d'oxygène des bras des ligands avec des acides aminés donneurs ou accepteurs d'hydrogène de la surface protéique.

3) L'interaction hydrophobe entre le macrocycle et des résidus aromatiques de la surface protéique.

4) L'établissement d'un réseau de liaisons hydrogènes à travers la coordination de molécules d'eau par les atomes d'oxygène (et d'azote) des bras des ligands. Ces molécules d'eau pouvant aussi établir des ponts vers la surface protéique.

5) L'interaction entre deux complexes menant à la fixation coopérative de deux molécules de complexe voisines.

Le tableau résume quelle sorte d'interaction décrite ci-dessus intervient lors de la fixation des complexes pour les différents dérivés. Avec pour abréviations : "**AfUOx**", urate oxydase d'*Aspergillus flavus* (Bonnete *et al.*, 2001), "**GI**", glucose isomérase de *Streptomyces rubiginosus* (Ramagopal *et al.*, 2003), "**TdTHAU**", thaumatine de *Thaumatococcus daniellii* (Ko *et al.*, 1994), "**HEWL**", lysozyme de blanc d'oeuf de poule (Girard *et al.*, 2003a), "**X**", protéine X, "**YGGV**", protéine hypothétique Yggv d'*Escherichia coli* (collaboration avec Chantal Abergel, IGS, Marseille), "**YEAZ**" : yeaZ d'*Escherichia coli* (Jeudy *et al.*, 2005).

	AfUOx	**GI**	**TdTHAU**	**HEWL**	**X**	**YGGV**	**YEAZ**
LnHPDO3A	-	-	-	3 4 5	2 3 4	-	-
LnDO3A	-	-	-	1 3 4 5	1 3	1	1 2
LnDOTMA^{1-}	2 3 4	-	-	-	4	2 4	1 4
LnDTPABMA	2 4	-	3	-	-	-	-
LnDTPA^{2-}		1 2 4	2 3 4	-	1 2 4	-	-
LnDOTA^{1-}	-	-	2 3	3 5	-	-	-
LnDOTA-BOM	-	-	2 3 4	2 3 4	-	-	-

Tableau 2.1 – Bilan des types de liaisons principales impliquées dans la fixation des complexes IRM testés, d'après Stelter (2005). Ce tableau ne récapitule que les modes de fixation ayant été déterminés au cours des travaux de recherche de Meike Stelter. (-) ne signifie pas que la protéine n'a pas pu être résolue à l'aide du complexe mais indique seulement que le mode de fixation n'a pas été déterminé.

2.1.3 La détection de la fixation des complexes IRM avant l'expérience de diffraction

Une des problématiques à laquelle l'équipe a toujours voulu répondre est : comment détecter la fixation des complexes avant l'étape de la diffraction ? L'idéal serait de pouvoir détecter la fixation des complexes en solution avant le processus de cristallisation afin de déterminer quel est le meilleur complexe à utiliser.

Dans ce but, Meike Stelter a, au cours de ces travaux de recherche, évalué l'utilisation de différentes méthodes physico-chimiques mais aucune n'a pu prétendre au titre de méthode facile à mettre en œuvre pour détecter la fixation des complexes : calorimétrie de titrage isotherme, résonance magnétique nucléaire 1 D et 2 D, résonance plasmonique de surface, électrophorèse de gels d'acrylamide natifs et expériences de luminescence en solution et dans les cristaux en utilisant le montage expérimental "Cryobench" (Ursby *et al.*, 2002). La plupart des méthodes ont été testées en utilisant le lysozyme de blanc d'œuf de poule comme protéine modèle et en travaillant, dans la mesure du possible, avec des conditions physico-chimiques proches des conditions de cristallisation.

La difficulté à laquelle elle a été confrontée pour établir une corrélation entre les observations en solution et celles dans le cristal peut s'expliquer par le fait que les complexes n'établissent que des liaisons non-covalentes avec des affinités faibles pour les protéines.

Jusqu'à présent, nous n'avons pas trouvé de méthode physico-chimique, fiable, aisée et peu coûteuse, permettant de détecter la fixation des complexes IRM. Mais un nouveau complexe exploitant les propriétés de luminescence des lanthanides a pu être synthétisé : le tris-dipicolinate de lanthanide. Sa luminescence dans le visible permet de détecter sa présence dans les cristaux.

2.2 Un complexe de lanthanide luminescent : le tris-dipicolinate d'europium

Dans le cas des complexes issus de l'imagerie médicale, au moins une molécule d'eau vient compléter la sphère de coordination du lanthanide, déplaçant la longueur d'onde de la fluorescence émise par le lanthanide hors du domaine du visible. Contrairement à ces complexes, dans le cas du complexe tris-dipicolinate de terbium ou d'europium, la luminescence intrinsèque de ces atomes de lanthanide est mesurable dans le visible.

L'idée de complexes luminescents permettant de détecter une éventuelle incorporation dans les cristaux avait déjà émergée lors des travaux de recherche de Marie-Laure Chesne. La première approche qu'elle avait entreprise s'appuyait alors sur des colorants utilisés dans l'industrie textile. Ces molécules "colorées" sont capables de fixer un ion uranyle, propriété intéressante puisque l'uranium présente une diffusion anomale encore plus importante que les lanthanides. Malheureusement, l'introdution de telles molécules dans les cristaux s'est avérée assez difficile, même si l'une d'entre elles a permis la détermination de la paroxonase humaine (Contreras-Martel *et al.*, 2006; Morales *et al.*, 2006).

Toutefois, cette approche a ouvert la voie à la recherche de nouvelles molécules facilement détectables, voie qui a conduit à la synthèse du complexe tris-dipicolinate de lanthanide, ou "$[Ln(DPA)_3]^{3-}$", qui fit l'objet des travaux de recherche de Guillaume Pompidor.

2.2.1 Présentation

Le complexe tris-dipicolinate de lanthanide est formé de trois ligands dissociés qui complètent la première sphère de coordination du lanthanide chélaté. Ces trois ligands sont des acides dipicoliniques (ou dipicolinate = pyridine-2,6-dicarboxylate = DPA). De par les deux fonctions acides que porte chaque ligand dipicolinate, ce complexe non-cagé est trois fois négativement chargé **[Figure 2.4]**. Son poids moléculaire théorique sans atome de lanthanide est de 636,34 $g.mol^{-1}$ car, en solution, trois contre-ions sodium viennent se fixer aux trois oxygènes de la fonction carboxylate de chaque ligand.

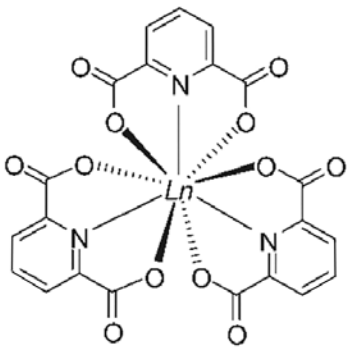


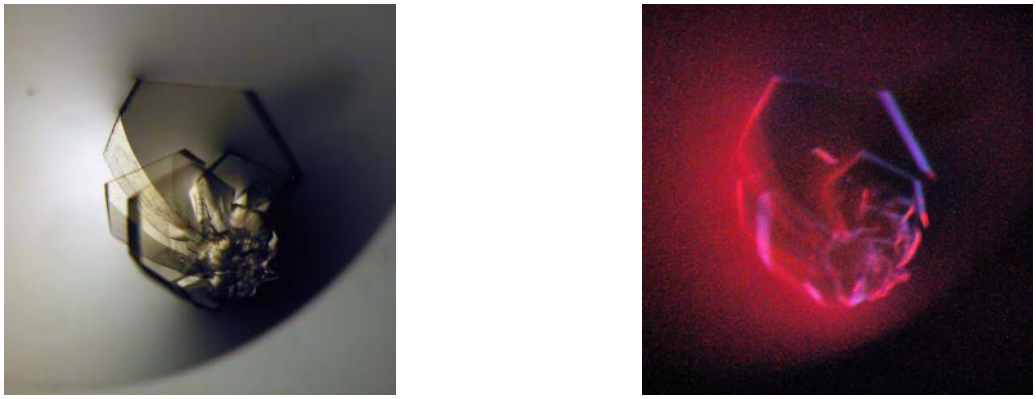
Figure 2.4 – Le tris-dipicolinate de lanthanide

Lorsque les ligands dipicolinate chélatent un ion lanthanide, dont le spectre d'absorption est dans le domaine des proches ultra-violets et dont le spectre d'émission de fluorescence est dans le domaine du visible, le complexe, excité sous UV proches, luminesce alors dans le visible. Comme nous l'avons indiqué dans le chapitre précédent, seuls deux lanthanides ont leur spectre d'émission de fluorescence dans le visible : le terbium et l'europium. Les complexes $[Tb(DPA)_3]^{3-}$ et $[Eu(DPA)_3]^{3-}$ luminescent respectivement dans le vert et le rouge.

Le complexe "$[Ln(DPA)_3]^{3-}$ présente une bonne solubilité dans l'eau (220 mM, à 25°C). Comme l'a indiqué Guillaume Pompidor, la présence de sels à fortes concentrations ou la présence de cations divalents conduisent à la dissociation du complexe. Un complexe tris-dipicolinate de zinc, par exemple, a été mis en évidence dans la structure de la thaumatine résolue par phasage expérimental à l'aide du complexe. Ce complexe ne peut donc être utilisé dans n'importe quelle condition de cristallisation mais, de par sa luminescence dans le visible, il est possible de détecter sa fixation dans les cristaux (*cf.* premier chapitre de la première partie).

2.2.2 La détection de la fixation du complexe $[Eu(DPA)_3]^{3-}$ avant l'expérience de diffraction

Les ligands dipicolinate jouant le rôle d'antenne, la luminescence intrinsèque des lanthanides terbium et europium dans le visible peut être facilement caractérisée.



(a) Cristal dérivé de lysozyme de blanc d'œuf de poule observé en lumière visible

(b) Cristal dérivé de lysozyme de blanc d'œuf de poule observé sous UV proche

Figure 2.5 – La luminescence du complexe $[Eu(DPA)_3]^{3-}$ dans un cristal dérivé de lysozyme de blanc d'œuf de poule

Comme le montre la **figure 2.5**, il est possible de mettre en évidence la fixation du complexe à l'aide d'une simple lampe UV de paillasse. Bien que cette méthode permette de révéler la fixation du complexe dans les cristaux, elle ne permet pas d'avoir une idée du taux de fixation dans le cristal. Cependant, Guillaume Pompidor a pu quantifier cette luminescence en menant une expérience de microscopie deux photons sur des cristaux dérivés de lysozyme avec les complexes $[Tb(DPA)_3]^{3-}$ et $[Eu(DPA)_3]^{3-}$ (D'Aléo *et al.*, 2007).

2.2.3 Etat de nos connaissances sur le mode de fixation du complexe $[Ln(DPA)_3]^{3-}$

Au cours des travaux de recherche de Guillaume Pompidor, les structures de six protéines ont été déterminées suite à un phasage expérimental s'appuyant sur le signal anomal du lanthanide introduit dans les cristaux dérivés grâce au complexe $[L(DPA)_3]^{3-}$: le lysozyme de blanc d'œuf de poule ("HEWL"), le lysozyme de blanc d'œuf de dinde, la thaumatine de Thaumatococcus daniellii, l'urate oxidase d'*Aspergillus flavus*, la xylanase de *Trichoderma reesei* et l'ubiquitine de coeur de bœuf.

Le mode de fixation du complexe $[Ln(DPA)_3]^{3-}$ a pu être déterminé de façon précise. Il se fixe aux arginines et aux lysines présentes dans la protéine, via des liaisons hydrogènes entre les groupements carboxylate des ligands dipicolinate et les groupements amine des acides aminés précités **[Figure 2.6]**. A la surface du HEWL, le complexe se fixe aussi en établissant une interaction hydrophobe de type empilement-π avec un acide aminé tryptophane. En co-cristallisation, les cristaux dérivés apparaissent généralement dans un précipité assez dense.

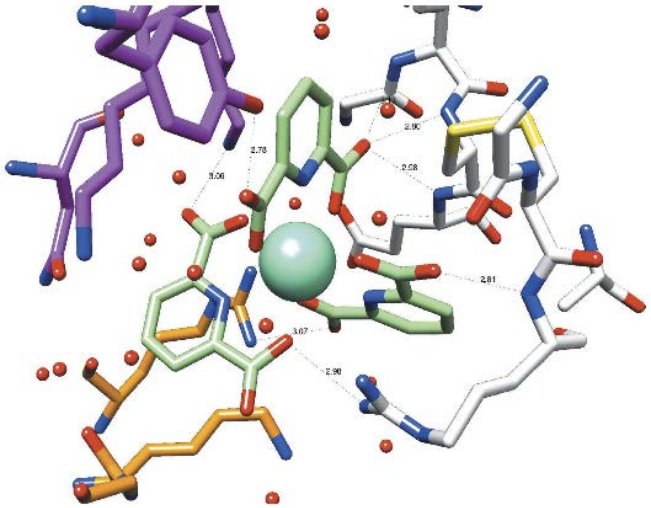


Figure 2.6 – Mode de fixation principal du complexe $[Ln(DPA)_3]^{3-}$ observé à la surface du HEWL (**?**)

Dans trois cas, le complexe $[Ln(DPA)_3]^{3-}$ a induit un effet supramoléculaire qui conduit à un empilement cristallin différent de celui observé pour la protéine native. Dans les conditions de cristallisation conduisant habituellement aux cristaux natifs et dans le cas des protéines :

1. lysozyme de blanc d'œuf de poule, les cristaux dérivés obtenus sont de symétrie cristalline C2 au lieu de $P4_32_12$,
2. xylanase de *Trichoderma reesei,* les cristaux dérivés obtenus sont de symétrie cristalline C2 au lieu de $P2_12_12_1$,
3. ubiquitine de cœur de bœuf, les cristaux dérivés obtenus sont de symétrie cristalline C222 au lieu de $P4_332$.

Dans le cas du HEWL, les cristaux de forme monoclinique ont présenté une diffraction à très haute résolution : 0,8 Å (au lieu de 1,4 Å pour les cristaux de symétrie cristalline quadratique). L'analyse de l'empilement cristallin montre que le complexe $[Ln(DPA)_3]^{3-}$ ponte plusieurs molécules de protéine, ce qui induit le changement de groupe d'espace **[Figure 2.7]**.

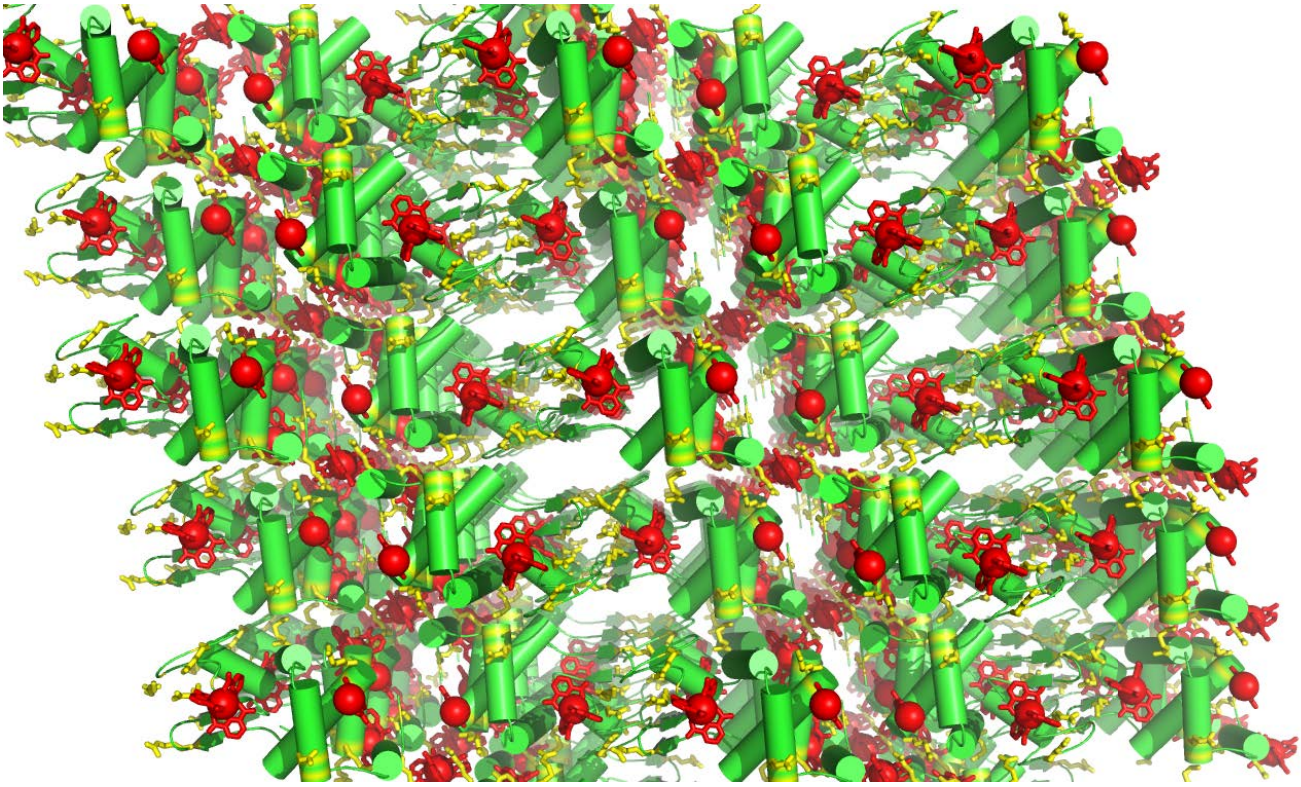


Figure 2.7 – Représentation de l'empilement cristallin observé dans les cristaux de HEWL de forme monoclinique obtenu en présence de $[Ln(DPA)_3]^{3-}$. Les molécules de complexe (en rouge) créent une interaction supramoléculaire avec les acides aminés arginine et lysine (en jaune) de la protéine (en vert).

De ces observations, nous avons envisagé l'utilisation du complexe $[Ln(DPA)_3]^{3-}$ en tant que complexe au fort pouvoir phasant capable de promouvoir la nucléation lorsqu'il est introduit dans la goutte en co-cristallisation. Cette hypothèse a été étudiée au travers d'une expérience de co-cristallisation automatisée décrite dans le premier chapitre de la première partie.

2.3 Protéines de structures inconnues déterminées jusqu'alors à l'aide des complexes de lanthanide

Lors de nos tests, nous évaluons le pouvoir phasant des complexes en déterminant la structure des protéines modèles comme s'il s'agissait de structures inconnues. Dès lors, dans la liste des structures inconnues déterminées à l'aide des complexes, on peut aussi inclure les protéines utilisées comme modèles d'étude. Ainsi, plus d'une quinzaine de structures de macromolécules biologiques ont été déterminées à l'aide des complexes de lanthanide. Parmi ces structures résolues, je m'attarde un peu plus sur la protéine dodécamérique chimérique Ornythile Transcarbamoyl 3630 (Girard *et al.*, 2003b) et la protéase octamérique Pab87 de *Pyrococcus horikoshii* (Delfosse *et al.*, 2009).

Ces deux assemblages protéiques, de par leur poids moléculaire supérieur à 400 kDa, illustrent le fort pouvoir phasant des complexes de lanthanide. Dans le cas de la protéine auto-compartimentée Pab87, de nombreux dérivés lourds ont d'abord été testés sans succès (sel de platine, d'iridium, d'osmium et d'holmium, protéine séléniée avec treize méthionines par monomère). C'est finalement des cristaux dérivés obtenus par co-cristallisation avec 200 mM de complexe LuHPDO3A qui ont permis le phasage

SAD à 2,0 Å de résolution des 3500 acides aminés compris dans l'unité asymétrique (groupe d'espace P1), et ce, sans recours à la symétrie non-cristallographique (Delfosse *et al.*, 2009).

Dans le cas de la protéine Pab87, 17 atomes de lutétium ont été modélisés à la surface de la protéine. Ceci prouve que le pouvoir phasant des lanthanides est suffisant pour le phasage d'un grand nombre d'acides aminés. De plus, le nombre de sites déterminés sur ces grands assemblages permet aussi d'émettre cette simple hypothèse : si les complexes de lanthanide ont trouvé des sites de fixation sur de petites protéines, la surface de protéines de bien plus grande taille devrait proposer davantage de sites de fixation adequat. C'est cette hypothèse qui a fait l'objet de l'étude décrite dans la seconde partie de ce manuscrit.

Deuxième partie

De nouveaux outils pour la détermination de la structure de macromolécules biologiques : la grande famille des complexes picolinates de lanthanide luminescents

Dans cette première partie méthodologique, je présente les différents développements entrepris autour du complexe tris-dipicolinate de lanthanide : $[Ln(DPA)_3]^{3-}$. Ce complexe, dont l'étude a fait l'objet de la thèse de Guillaume Pompidor (2004 - 2007), arbore des propriétés physico-chimiques qui le distingue des complexes issus de l'imagerie médicale. En effet, au cours de sa thèse, Guillaume Pompidor a montré que ce complexe a l'avantage de présenter une fluorescence dans le visible lorsqu'il est exposé au UV proches. Cette propriété permet de révéler la fixation du complexe dans les cristaux dérivés de protéine avant l'étape de diffraction. De plus, l'obtention de cristaux dérivés par co-cristallisation et l'étude du mode de fixation du complexe $[Ln(DPA)_3]^{3-}$ ont révélé une autre propriété qui lui est propre. Dans les conditions de cristallisation natives de deux protéines de petite taille, l'introduction de ce complexe a conduit à des cristaux d'un groupe d'espace différent de celui des cristaux natifs. En établissant des liaisons hydrogènes et hydrophobes (de type empilement-π) simultanément avec plusieurs molécules, ce complexe peut conduire à un empilement cristallin différent, ce que nous avons qualifié d' "effet supramoléculaire du tris-dipicolinate de lanthanide" (*cf. introduction*). Ainsi, ces deux propriétés font du complexe $[Ln(DPA)_3]^{3-}$ un complexe très prometteur : un complexe dont la fixation peut être facilement révélée et qui peut jouer le rôle d'agent de nucléation pour promouvoir l'apparition de cristaux dérivés en co-cristallisation.

C'est ce que nous avons voulu vérifié en premier lieu. Nous avons mené une expérience de co-cristallisation automatisée à l'aide du robot de cristallisation nanogoutte de l'EMBL avec six protéines éclectiques. Cette expérience nous a, en outre, permis de mettre en évidence et de façon statistique un défaut inhérent à l'utilisation du complexe $[Ln(DPA)_3]^{3-}$: son auto-cristallisation dans certaines conditions de cristallisation. Ce défaut avait déjà été souligné par Guillaume Pompidor. Les observations extraites de cette expérience font l'objet du premier chapitre.

Au cours de mes travaux de thèse, nous nous sommes impliqués dans l'élaboration de deux nouveaux types de complexes de lanthanide, tous deux dérivés du tris-dipicolinate de lanthanide. Notre implication dans ce développement a bien entendu été cristallographique.

Dans un second chapitre, je présente donc la nouvelle famille des complexes tris-triazoledipicolinates de lanthanide, "$[Ln(TDPA)_3]^{3-}$". Comme l'appellation le laisse entendre, il s'agit de complexes constitués de ligands dipicolinate auxquels ont été fixés, par chimie-clic, un noyau triazole portant un groupement chimique supplémentaire qu'il est possible de faire varier. Cette modification structurale apporte aux nouveaux complexes tris-triazoledipicolinates de lanthanide une plus forte affinité pour les macromolécules biologiques et donc un fort pouvoir de phasage, une luminescence accrue dans le visible et enfin, une versatilité dans leur mode de fixation aux protéines modèles. Cette versatilité est due à ce groupement chimique supplémentaire qu'il est possible de faire varier. Et c'est cette variabilité qui permet de faire ainsi un choix parmi différents ligands pour composer un complexe tris-triazoledipicolinate personnalisé et bâtir ainsi toute une famille de complexes. Mais "comme rien n'est parfait dans le meilleur des mondes", cette famille de complexes conserve le défaut d'utilisation du tris-dipicolinate de lanthanide d'origine : l'auto-cristallisation dans certaines conditions de cristallisation.

C'est pourquoi nous nous sommes intéressés à un nouveau complexe, le triazonanetripicolinate de lanthanide ou "LnTNTPA". A la différence du complexe $[Ln(DPA)_3]^{3-}$ et de ceux de la famille des "$[Ln(TDPA)_3]^{3-}$, ce complexe est cagé et sa charge globale est nulle. Un macrocycle triazacylononane relie les trois acides picoliniques et le tout forme une cage qui chélate l'ion lanthanide. D'une part, cette structure évite au complexe d'être déstabilisé par la présence d'ions divalents, une des conditions

qui engendre l'auto-cristallisation des autres complexes précités. D'autre part, elle prévient l'éventuelle substitution de l'ion lanthanide par un ion divalent au sein du complexe. Ce phénomène avait été observé avec le complexe $[Ln(DPA)_3]^{3-}$ et un ion Zn^{2+} dans la structure de l'ubiquitine de cœur de bœuf. En outre, le complexe conserve toutes les propriétés de luminescence. Les premières observations et les premiers résultats obtenus avec le complexe TbTNTPA sont consignés dans le troisième chapitre.

Chapitre 3

Le tris-dipicolinate de lanthanide, agent de nucléation ?

Au cours des travaux de recherche, Guillaume Pompidor a mis en évidence "l'effet supramoléculaire" du complexe luminescent tris-dipicolinate de lanthanide, $[Ln(DPA)_3]^{3-}$. Dans les conditions de cristallisation conduisant habituellement aux cristaux natifs, le complexe $Ln(DPA)_3]^{3-}$, introduit par co-cristallisation, a conduit à un empilement cristallin différent de celui habituellement observé pour trois protéines natives : le lysozyme de blanc d'œuf de poule (Pompidor *et al.*, 2008), la xylanase de *Trichoderma reesei,* et l'ubiquitine de cœur de bœuf.

Nous nous sommes alors demandé si, au delà de son utilisation en tant que diffuseur anomal, on pouvait utiliser le tris-dipicolinate de lanthanide en tant que petite molécule pouvant favoriser la formation de contacts cristallins et ainsi promouvoir la nucléation. L'ajout de petites molécules dans la goutte de cristallisation pour provoquer la nucléation et améliorer la croissance cristalline avait été proposé en 2006 par Mc Pherson et Cudney (McPherson and Cudney, 2006). Pour ce faire, les petites molécules créent des interactions faibles avec les protéines, comme les liaisons hydrogènes, les interactions électrostatiques ou hydrophobes. Ils ont aussi mis en évidence que certaines catégories de petites molécules sont particulièrement propices pour établir ces interactions : les acides dicarboxyliques (fonction chimique que porte le complexe $Ln(DPA)_3]^{3-}$) et les groupements diamines. Pour certaines protéines, l'ajout de ce type de molécules dans les gouttes de cristallisation conduit à l'apparition de différentes formes cristallines. Dans certains cas, des effets notables sur la qualité de diffraction des cristaux (augmentation de la résolution, baisse de la mosaïcité) ont été observés (Larson *et al.*, 2007).

Par ailleurs, comme l'avait déjà indiqué Guillaume Pompidor, la présence de fortes concentrations en sels ou la présence de cations divalents dans les conditions de cristallisation semble conduire à la dissociation du complexe. Je me suis moi-même confronté à la formation de cristaux de sel de complexe lors de la préparation de cristaux dérivés. En effet, dès que la solution de cristallisation contenait des ions divalents (Mg^{2+}, Ca^{2+}, Cd^{2+}) ou une concentration trop importante de sel, le complexe cristallisait dans la goutte.

C'est donc avec un double objectif que j'ai mené une expérience de co-cristallisation automatisée à l'aide du robot de cristallisation nanogoutte du laboratoire HTX de l'EMBL : vérifier d'une part l'utilisation de ce complexe en tant qu'agent facilitant la nucléation et, d'autre part, essayer de mieux comprendre ce problème d'auto-cristallisation.

3.1 Mise en place de l'expérience de co-cristallisation automatisée

3.1.1 Choix des protéines modèles

Six protéines différentes ont été choisies pour mener à bien cette expérience :

1. L'Acriflavine resistance protein B de la bactérie *Escherichia coli*, "EcAcrB", protéine membranaire homotrimèrique de 341 kDa, produite et purifiée par mes soins (cf. annexes),
2. L'Enhanced Green Fluorescent Protein de la méduse *Aequorea victoria*, "AvEGFP", protéine cytosolique monomèrique de 24 kDa, produite et purifiée par mes soins (cf. annexes),
3. La métalloprotéine CnrXs de la souche CH34 de la bactérie *Cupriavidus metallidurans*, "CmCH34CnrXs", protéine périplasmique de 26 kDa, produite et purifiée par le Dr. Juliette Trepreau de l'équipe Métalloprotéines de l'institut (Pompidor *et al.*, 2009),
4. La malate déshydrogénase de la bactérie thermophile *Chloroflexus aurantiacus,* "CaMalDH", protéine cytosolique homotétramérique de 131 kDa, produite et purifiée par le Dr. Dominique Madern, membre de notre groupe E.L.M.A. (cf. second chapitre de la deuxième partie),
5. L'urate oxydase du champignon *Aspergillus flavus,* "AfUox", protéine cytosolique homotétramèrique de 135 kDa, fournie par le Dr. Nathalie Colloc'h de l'équipe CERVOxy du laboratoire Imagerie et Stratégie Thérapeutiques des pathologies Cérébrales et Tumorales, à Caen et par le Pr. Thierry Prangé du Laboratoire de Cristallographie et RMN Biologiques de l'Université Paris V René Descartes, à Paris (Bonnete *et al.*, 2001),
6. L'aminopeptidase TET2-12s de l'archaea *Pyrococcus horikoshii,* "PhTET2-12s", protéine homododécamérique tétraédrique de 450 kDa produite et purifiée par la Dr. Eva Rosenbaum et Mylène Ferruit, membre de notre groupe E.L.M.A.

Notre choix fut orienté par plusieurs critères. Nous avons choisi un éventail de protéines présentant différentes tailles, différentes propriétés de surface (charges, cavités...) et différentes propriétés cristallines (proportion de solvant, empilement). Nous avons travaillé avec des protéines pour lesquelles des données cristallographiques sont disponibles (AvEGFP, EcAcrB, CmCH34CnrXs). En effet, pour ce type d'expérience où l'examen minutieux de chaque faciès cristallin pour chaque condition de cristallisation est nécessaire, il est judicieux de travailler avec des protéines dont les conditions de cristallisation et les formes cristallines associées sont connues. Cela permet ainsi de dépister les nouveaux faciès cristallin dus à la présence du $[Ln(DPA)_3]^{3-}$ dans la goutte. Nous avons utilisé les protéines produites et purifiées au laboratoire en quantité suffisante pour une telle expérience (AvEGFP, EcAcrB). Nous avons aussi profité de celles fournies par nos collaborateurs pour lesquelles nous recherchions de nouvelles conditions de cristallisation (AfUox, CaMalDH, PhTET2-12s et CmCH34CnrXs).

3.1.2 Choix des conditions de co-cristallisation

Les robots de cristallisation nanogoutte sont de plus en plus utilisés pour déterminer rapidement d'éventuelles conditions de cristallisation. Pour cette expérience, afin qu'elle soit le reflet d'une expérience ordinaire de cristallisation, nous avons choisi les six kits de cristallisation commerciaux **[Tableau 3.1]** et le stockage des plaques de cristallisation à température ambiante. 576 conditions de cristallisation par protéine ont ainsi pu être analysées.

Plaque	Nom des kits Hampton®	Noms des kits équivalent Qiagen
1	Crystal screen I & II	The Classics
2	Crystal screen lite & PEG/Ion	Mélange HTX pour plaque n°2
3	MembFac & Natrix	Mélange HTX pour plaque n°3
4	Quick screen & Grid screens sulfate, formate, malonate	Mélange HTX pour plaque n°4
5	Grid screens PEG 6K, PEG/LiCl, MPD - Screen Mme	Mélange HTX pour plaque n°5
6	Index screen	The Classics II

Tableau 3.1 – Kits commerciaux de cristallisation utilisés au robot HTX

3.1.3 Matériel & méthodes

3.1.3.1 Composition des échantillons

Pour les protéines EcAcrB, AvEGFP, CmCH34CnrXs et CaMalDH, deux échantillons ont été apportés au robot HTX : l'un ne contenant que la protéine concentrée et l'autre contenant la protéine et le complexe $[Ln(DPA)_3]^{3-}$. Dans le cas des protéines AfUox et PhTET2-12s, un troisième échantillon, faisant office de contrôle supplémentaire et ne contenant que le $[Ln(DPA)_3]^{3-}$, a été analysé au robot. Ce troisième échantillon m'a permis, d'une part, d'affiner mes observations quant au comportement du complexe en présence et en absence de protéine, et d'autre part, de me servir de contrôle visuel pour la comparaison des faciès cristallin.

Le robot de cristallisation nanogoutte du laboratoire HTX à l'EMBL, Grenoble, utilise des plaques de cristallisation 96 puits Greiner® avec 3 emplacements de gouttes assises par puits. Hormis AcrB qui a été cristallisée en goutte suspendue sur le film plastique qui recouvre hermétiquement la plaque, (la présence de détergent provoquant sinon l'étalement de la goutte dans les emplacements pour goutte assise) toutes les autres protéines ont été cristallisées selon la méthode de la goutte assise. Le robot dépose 100 µL de solutions de cristallisation qui diffèrent dans chacun des 96 puits. Puis, pour former la goutte, il prélève et dépose 100 nL d'un échantillon avant d'y ajouter 100 nL de puits. La présence de trois emplacements par puits m'a permis, pour chaque protéine, de synchroniser la cristallisation de la protéine seule et la co-cristallisation de la protéine avec le complexe. Le volume d'un échantillon requis pour une expérience de cristallisation avec ces six kits est de 90 µL.

Le **tableau 3.2** récapitule les concentrations en protéine en en complexe dans la goutte de cristallisation formée par le robot. Toutes les valeurs des concentrations indiquées correspondent à celles calculées "à t_0", c'est-à-dire au moment où la goutte de cristallisation est formée. Pour retrouver la valeur des concentrations respectives des échantillons apportés au laboratoire HTX, il faut donc multiplier les valeurs indiquées par le facteur de dilution de 2. Pour les gouttes contenant à la fois la protéine et le complexe $[Ln(DPA)_3]^{3-}$, nous avons fait en sorte que la concentration en complexe "à t_0" soit la même que celle lors d'une expérience de co-cristallisation en laboratoire. La solubilité du complexe $[Ln(DPA)_3]^{3-}$ à température ambiante étant de 220 mM, lorsque nous ajoutons un volume de solution de complexe sur une goutte formée par un même volume de liqueur mère et un même volume de solution de protéine, le facteur de dilution est, à t_0, de 3. Dans la goutte, la concentration est donc d'environ 73 mM. Ce **tableau 3.2** indique aussi la composition du tampon dans lequel chaque protéine était stockée avant l'experience, ceci pour eventuellement correler les résultats de cristallisation avec cette compostion, si cela s'averera necessaire.

Experience	Goutte	Concentrations dans la goutte à t_0	Tampon de stockage de la protéine
EcAcrB	1	4 mg.mL^{-1} EcAcrB	10 mM HEPES pH 7,0 ; 0,02 % ß-DDM
	2	4 mg.mL^{-1}EcAcrB + 73 mM $[Eu(DPA)_3]^{3-}$	
AvEGFP	1	15 mg.mL^{-1} AvEGFP	20 mM HEPES pH 7,5
	2	15 mg.mL^{-1} AvEGFP + 73 mM $[Eu(DPA)_3]^{3-}$	
CmCH34CnrXs	1	10 mg.mL^{-1} CmCH34CnrXs	20 mM HEPES pH 7,5 ; 120 mM NaCl
	2	10 mg.mL^{-1} CmCH34CnrXs + 73 mM $[Eu(DPA)_3]^{3-}$	
CaMalDH	1	10 mg.mL^{-1} CaMalDH	50 mM TRIS HCl pH 8,0
	2	10 mg.mL^{-1} CaMalDH + 73 mM $[Eu(DPA)_3]^{3-}$	
AfUox	1	10 mg.mL^{-1} AfUox	100 mM TRIS HCl pH 8.0 ; 50 mM NaCl
	2	10 mg.mL^{-1} AfUox + 73 mM $[Eu(DPA)_3]^{3-}$	
	3	25 mM $[Eu(DPA)_3]^{3-}$	
PhTET2-12s	1	8 mg.mL^{-1} PhTET2-12s	50 mM TRIS HCl pH 7,5 ; 150 mM NaCl
	2	8 mg.mL^{-1} PhTET2-12s + 73 mM $[Eu(DPA)_3]^{3-}$	
	3	100 mM $[Eu(DPA)_3]^{3-}$	

Tableau 3.2 – Concentrations en protéine et en complexe dans chaque goutte de cristallisation formée par le robot HTX. Le tampon de stockage initial dans lequel la protéine était conservée est également indiqué.

3.1.3.2 Observations des plaques

Les plaques de cristallisation réalisées ont été placées dans le "dispositif de photographies automatisé" RoboDesign Minstrel III (Rigaku®) à 20°C. Après dépôt, chaque goutte a été successivement photographiée six fois selon les intervalles suivants : 1 jour, 3 jours, 7 jours, 16 jours, 5 semaines et 10 semaines. Ces photos sont consultables en ligne et m'ont permis de suivre l'évolution de chacune des gouttes de cristallisation.

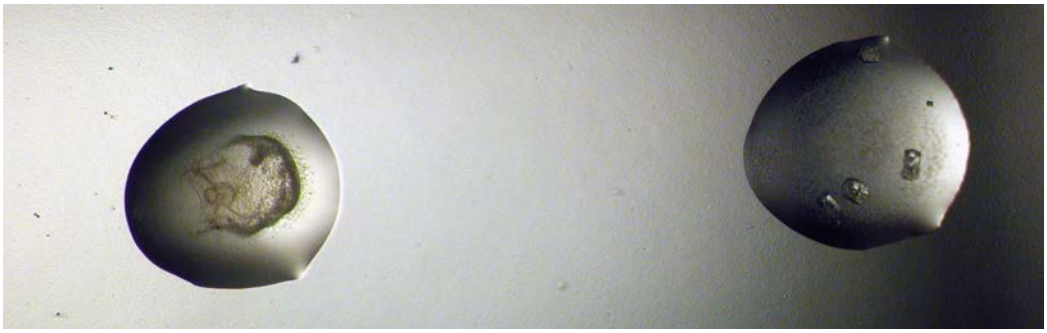
3.1.3.3 Contrôle de la luminescence dans les gouttes

Une fois les 10 semaines écoulées, les 36 plaques ont été récupérées et observées une nouvelle fois mais sous binoculaire. Cette dernière étape m'a permis de mettre les plaques sous exposition aux UV proches (λ = 315 nm), à l'aide d'une lampe à UV de paillasse, afin de vérifier et d'évaluer approximativement la luminescence des cristaux présents dans les gouttes.

Cette étape m'a permis de discriminer les "*faux positifs*" des "*vrais positifs*" :

- ▷ un "*faux positif*" est un cristal potentiellement dérivé, c'est-à-dire ayant crû en présence de $[Ln(DPA)_3]^{3-}$ et présentant un faciès différent de celui d'un cristal natif, mais qui se révèle être en fait un cristal de sel de $[Ln(DPA)_3]^{3-}$,
- ▷ un "*vrai positif*" est un cristal ayant poussé en présence de $[Ln(DPA)_3]^{3-}$, présentant un faciès différent de celui d'un cristal natif et qui se révèle être un cristal dérivé.

Contrôler la luminescence des cristaux, même très approximativement, permet en effet de discriminer ces deux types de cristaux car les cristaux de sel de $[Eu(DPA)]^{3-}$ présentent une luminescence bien plus forte dans le visible que les cristaux dérivés **[Figure 3.1]**. L'exemple ci-dessous illustre ce contrôle et les "fausse joies" auxquelles nous avons du faire face. Dans le cadre de l'expérience menée avec la protéine EcAcrB, la goutte contenant le complexe (à droite) pouvait laisser présager de cristaux dérivés. Malheureusement, aux vues de leur luminescence, ils se sont avérés être des cristaux de sel de $[Eu(DPA)]^{3-}$ **[Figure 3.1].**



(a) Goutte de gauche : EcAcrB, goutte de droite : EcAcrB + $[Eu(DPA)_3]^{3-}$, lumière visible

(b) Goutte de gauche : EcAcrB, goutte de droite : EcAcrB + $[Eu(DPA)_3]^{3-}$, UV proches

Figure 3.1 – Détection de cristaux de sels de $[Eu(DPA)_3]^{3-}$ sous UV proches, dans le cadre de l'expérience menée avec EcAcrB

Avant cela, lors de l'observation des photos en ligne, j'ai aussi pu m'aider de trois autres critères pour discriminer les "*faux négatifs*" des "*vrais positifs*". Le complexe est très sensible à l'évaporation, ce qui provoque sa cristallisation immédiate en bordure de goutte. De plus, les cristaux de sel de $[Ln(DPA)_3]^{3-}$ apparaissent très rapidement, en une nuit, avant la première observation de la plaque par le dispositif de photographies automatisé. Enfin, dans le cas des protéines AfUox et PhTET2-12s, l'observation de la troisième goutte contenant uniquement le complexe s'est révélé un excellent contrôle. Il m'a été alors possible de comparer le profil des cristaux dans les trois gouttes pour mieux différencier les cristaux de sel de $[Ln(DPA)_3]^{3-}$ des éventuels cristaux dérivés.

3.1.4 La démarche retenue pour l'analyse statistique des expériences de co-cristallisation

Pour faciliter la compréhension des résultats qui vont être présentés, nous décrivons ici la démarche qui a été retenue pour ne retenir que les observations pertinentes pour cette étude.

Cette étude nécessitant l'observation et l'analyse comparative de $(2 \times 576) \times 4$ gouttes (pour les quatre premières protéines précitées) et de $(3 \times 576) \times 2$ gouttes (pour les protéines AfUox et PhTET1-12s), nous avons ciblé des profils de goutte afin de ne pas se fourvoyer dans une étude comparative fastidieuse.

Avec deux gouttes déposées par puits, la première ne contenant que la protéine, la seconde contenant la protéine avec le complexe, voici la liste exhaustive des profils de goutte observables :

Goutte n°1, protéine seule :

- ▷ goutte claire,
- ▷ précipité protéique,
- ▷ cristaux natifs.

Goutte n°2, protéine et $[Eu(DPA)_3]^{3-}$:

- ▷ goutte claire,
- ▷ précipité protéique,
- ▷ précipité protéique + cristaux de $[Eu(DPA)_3]^{3-}$,
- ▷ cristaux natifs,
- ▷ cristaux dérivés,
- ▷ cristaux de $[Eu(DPA)_3]^{3-}$,
- ▷ cristaux natifs + cristaux de $[Eu(DPA)_3]^{3-}$,
- ▷ cristaux natifs + cristaux dérivés + cristaux de $[Eu(DPA)_3]^{3-}$,
- ▷ cristaux dérivés + cristaux natifs,
- ▷ cristaux dérivés + cristaux de $[Eu(DPA)_3]^{3-}$.

L'observation des gouttes, deux à deux, conduisait à un trop grand nombre de combinaisons possibles et toutes n'étaient pas utiles pour cette étude. C'est ainsi que, après avoir examiné un peu moins de la moitié des plaques, j'ai pu réduire le nombre de combinaisons pour mieux classer mes observations.

Premièrement, je suis parti du principe que l'observation d'un "précipité protéique" en sus de cristaux natifs dans la goutte est tout à fait banal. La catégorie "précipité protéique" n'avait donc pas lieu d'être ajoutée et mentionnée dans ce cas. Deuxièmement, je n'ai jamais rencontré : ni de précipité de $[Eu(DPA)_3]^{3-}$, ni de présence conjointe de cristaux natifs et de cristaux dérivés, ni de présence conjointe de cristaux dérivés et de cristaux de sel de $[Eu(DPA)_3]^{3-}$ et encore moins de cristaux natifs, de cristaux dérivés et de cristaux de sel de $[Eu(DPA)_3]^{3-}$ dans une même goutte. Le nombre de conditions qui ne purent être étudiées à cause d'un défaut dans l'élaboration de la goutte par le robot a aussi été évalue, ceci permettant de normaliser les proportions de chaque catégorie vis-à-vis du nombre total de conditions observables. J'ai donc pu, pour classer mes observations, réduire le nombre de catégories à sept :

- ▷ goutte claire
- ▷ conditions menant à l'apparition de cristaux dérivés potentiels,
- ▷ conditions menant à l'apparition de cristaux natifs,
- ▷ conditions menant à l'apparition de cristaux de sel de $[Eu(DPA)_3]^{3-}$ en présence de protéine,
- ▷ conditions dans lesquelles la présence de complexe provoque la précipitation de la protéine ou la formation de cristaux natifs,
- ▷ conditions dans lesquelles la présence de complexe provoque la solubilisation de la protéine (disparition du précipité protéique ou dissolution des cristaux natifs).
- ▷ défaut du robot

Dans le cas des protéines AfUox et PhTET2-12s, la troisième goutte contenant uniquement le complexe a permis d'ajouter des précisions concernant la quatrième catégorie précitée (l'apparition de cristaux de sel de $[Eu(DPA)_3]^{3-}$) :

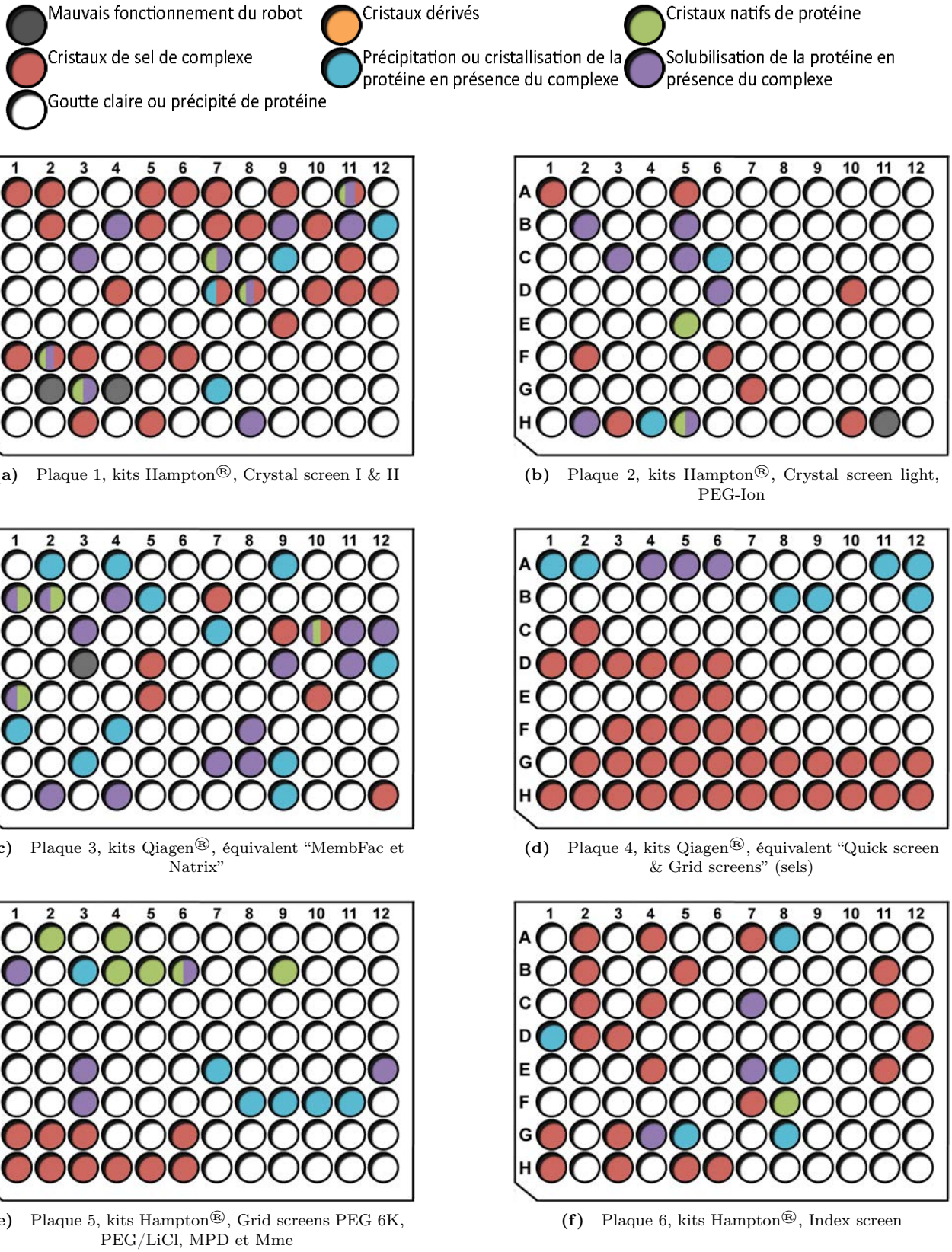
- ▷ conditions dans lesquelles le complexe ne cristallise qu'en présence de protéine,
- ▷ conditions dans lesquelles le complexe ne cristallise qu'en absence de protéine,
- ▷ conditions dans lesquelles le complexe cristallise, qu'il soit en présence de protéine ou non.

Enfin, le fait de déposer une troisième goutte contenant uniquement le complexe et ce, à deux concentrations différentes (25 mM et 100 mM), nous a permis de mesurer l'influence de sa concentration sur sa cristallisation en absence de protéine.

3.2 Les observations issues de l'expérience de co-cristallisation automatisée avec le $[Eu(DPA)_3]^{3-}$

Nous rapportons, dans cette partie, l'ensemble des observations recueillies lors des l'expériences de co-cristallisation automatisée menées avec chacune des protéines. Afin de ne pas alourdir le manuscrit, je ne donne qu'un exemple illustrant la façon dont j'ai procédé pour noter l'ensemble des observations **[figure 3.2]**.

3.2.1 Co-cristallisation de la protéine EcAcrB avec le complexe $[Eu(DPA)_3]^{3-}$


Mauvais fonctionnement du robot
Cristaux dérivés
Cristaux natifs de protéine
Cristaux de sel de complexe
Précipitation ou cristallisation de la protéine en présence du complexe
Solubilisation de la protéine en présence du complexe
Goutte claire ou précipité de protéine
1 2 3 4 5 6 7 8 9 10 11 12
A B C D E F G H

(a) Plaque 1, kits Hampton®, Crystal screen I & II

(b) Plaque 2, kits Hampton®, Crystal screen light, PEG-Ion

(c) Plaque 3, kits Qiagen®, équivalent "MembFac et Natrix"

(d) Plaque 4, kits Qiagen®, équivalent "Quick screen & Grid screens" (sels)

(e) Plaque 5, kits Hampton®, Grid screens PEG 6K, PEG/LiCl, MPD et Mme

(f) Plaque 6, kits Hampton®, Index screen

Figure 3.2 – Observations des plaques de co-cristallisation EcAcrB avec le complexe $[Eu(DPA)_3]^{3-}$

Les premières observations issues de l'expérience de co-cristallisation menée avec la protéine AcrB confirment celles de Guillaume Pompidor : quel que soit le type de sel, dès que sa concentration dépasse 0,8 M, on observe une auto-cristallisation du tris-dipicolinate d'europium dans la goutte. La plaque de cristallisation 4 qui contient les liqueurs mères à base de hautes concentrations en sels provoque l'apparition de cristaux de sel de $[Eu(DPA)_3]^{3-}$ dans plus de 38 % des gouttes **[Figure 3.3]**.

Sur l'ensemble des plaques de cristallisation, le complexe cristallise dans 19 % des conditions de cristallisation. Nous n'avons pas constaté de goutte où des cristaux dérivés étaient apparus.

En outre, si on ne se concentre que sur les observations relatives à l'influence du tris-dipicolinate d'europium sur la cristallisation en goutte suspendue, on met en évidence le fait que son auto-cristallisation est sa manifestation la plus fréquente **[Figure 3.4]**.

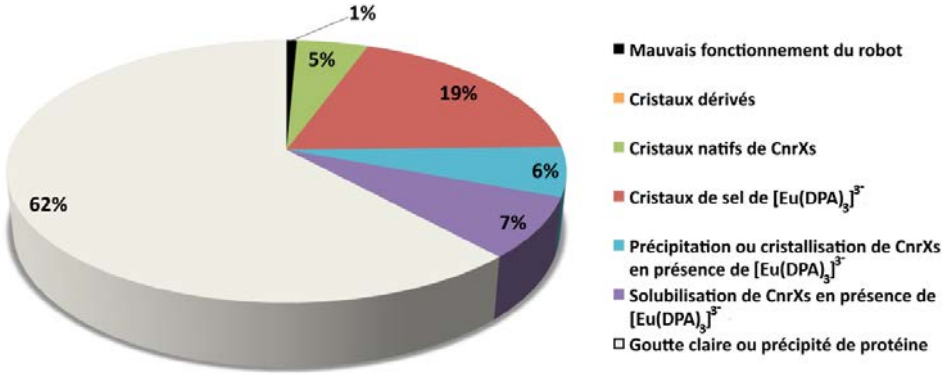


Figure 3.3 – Bilan statistique des observations des essais de co-cristallisation de EcAcrB avec le complexe $[Eu(DPA)_3]^{3-}$

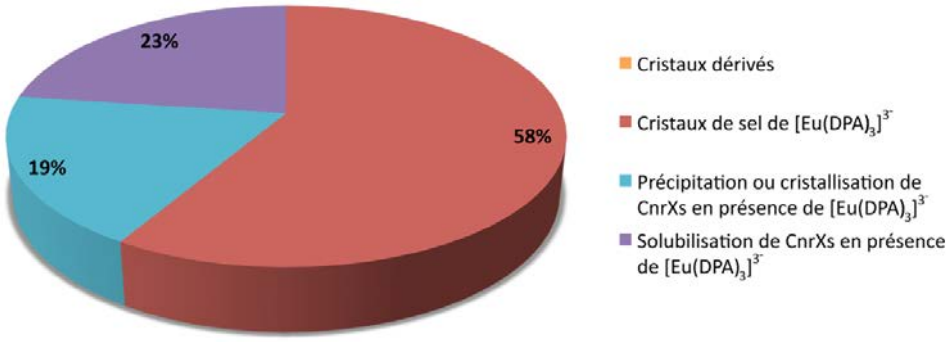


Figure 3.4 – Bilan statistique des observations suite à l'ajout du complexe $[Eu(DPA)_3]^{3-}$ en co-cristallisation avec EcAcrB

3.2.2 Co-cristallisation de la protéine AvEGFP avec le complexe $[Eu(DPA)_3]^{3-}$

Les même essais de co-cristallisation ont été menés avec la protéine AvEGFP.

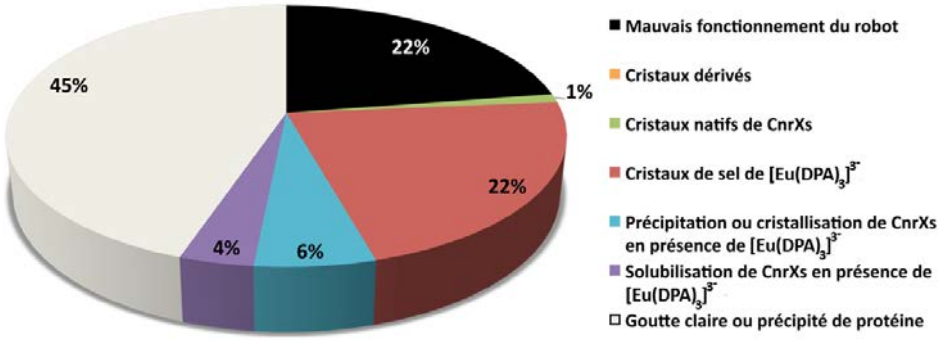


Figure 3.5 – Bilan statistique des observations des essais de co-cristallisation de AvEGFP avec le complexe $[Eu(DPA)_3]^{3-}$

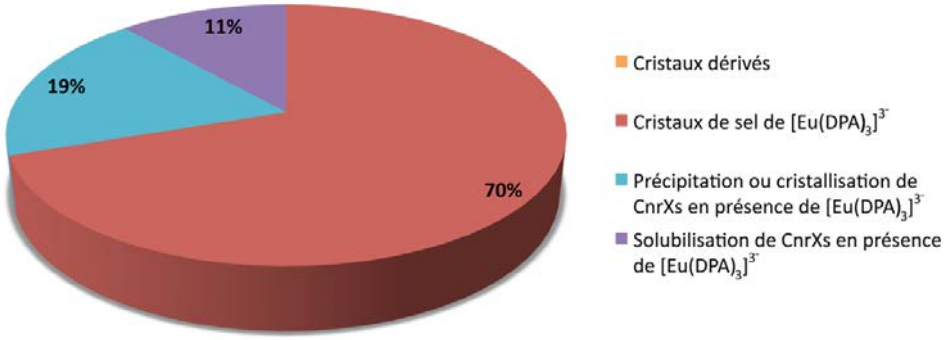


Figure 3.6 – Bilan statistique des observations suite à l'ajout du complexe $[Eu(DPA)_3]^{3-}$ en co-cristallisation avec AvEGFP

En examinant les gouttes de la même manière que pour la protéine EcAcrB et en conservant le même code couleur, nous voyons sur les **figures 3.5** et **3.6** qu'il n'a pas été possible d'obtenir des cristaux dérivés de AvEGFP et que de toutes les observations, l'auto-cristallisation du complexe $[Eu(DPA)_3]^{3-}$ est la plus courante.

3.2.3 Co-cristallisation de la protéine CmCH34CnrXs avec le complexe $[Eu(DPA)_3]^{3-}$

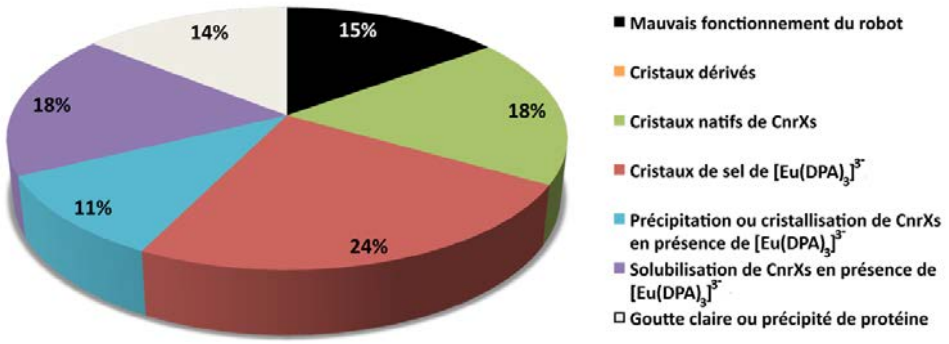


Figure 3.7 – Bilan statistique des observations des essais de co-cristallisation de CmCH34CnrXs avec le complexe $[Eu(DPA)_3]^{3-}$

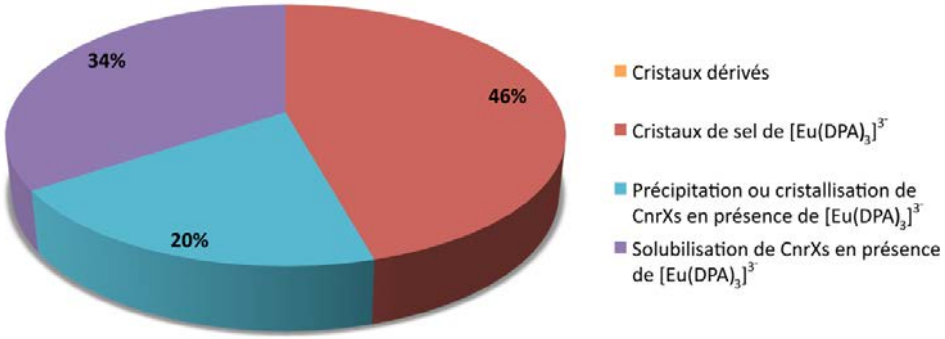


Figure 3.8 – Bilan statistique des observations suite à l'ajout du complexe $[Eu(DPA)_3]^{3-}$ en co-cristallisation avec CmCH34CnrXs

L'examen des deux gouttes dans les 576 conditions de cristallisation a conduit au même constat : le complexe $[Eu(DPA)_3]^{3-}$ a auto-cristallisé dans 24 % des conditions utilisées, ce qui correspond à nouveau à la part la plus importante des observations et correspond à 46 % de l'ensemble des profils retenus pour décrire l'influence du complexe en co-cristallisation **[Figure 3.7 et 3.8]**.

3.2.4 Co-cristallisation de la protéine CaMalDH avec le complexe $[Eu(DPA)_3]^{3-}$

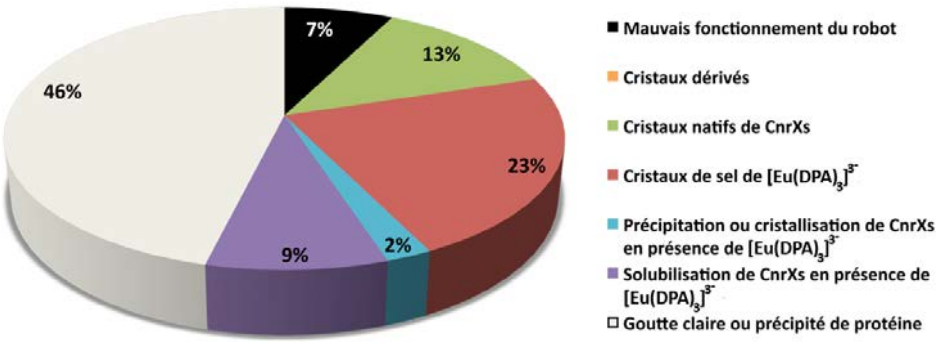


Figure 3.9 – Bilan statistique des observations des essais de co-cristallisation de CaMalDH avec le complexe $[Eu(DPA)_3]^{3-}$

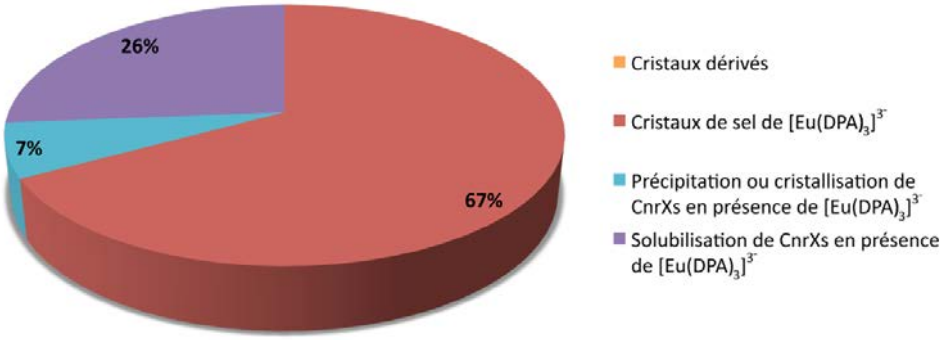


Figure 3.10 – Bilan statistique des observations suite à l'ajout du complexe $[Eu(DPA)_3]^{3-}$ en co-cristallisation avec CaMalDH

Encore une fois, la **figure 3.9** indique que l'apparition de cristaux de sel de $[Eu(DPA)_3]^{3-}$ survient dans 23 % des conditions de cristallisation. Si on s'en tient uniquement aux observations des gouttes où le complexe est présent, 67 % de celles-ci sont concernées par l'apparition de ces cristaux de sel de complexe **[Figure 3.10]**.

3.2.5 Co-cristallisation de la protéine AfUox avec le complexe $[Eu(DPA)_3]^{3-}$

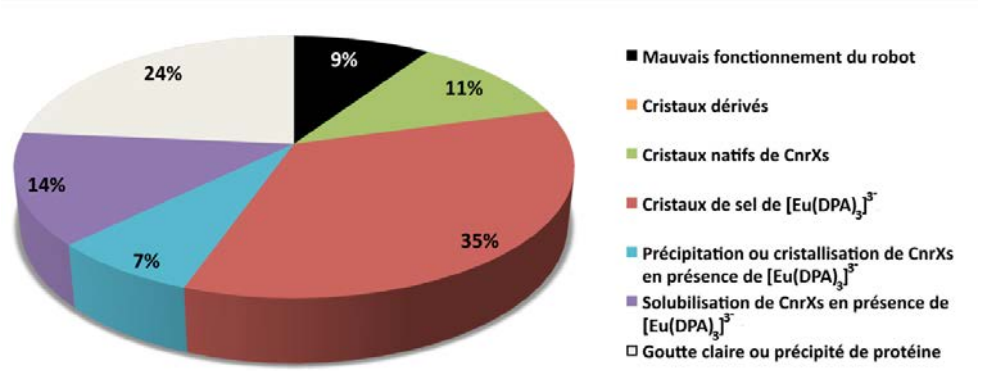


Figure 3.11 – Bilan statistique des observations des essais de co-cristallisation de AfUox avec le complexe $[Eu(DPA)_3]^{3-}$

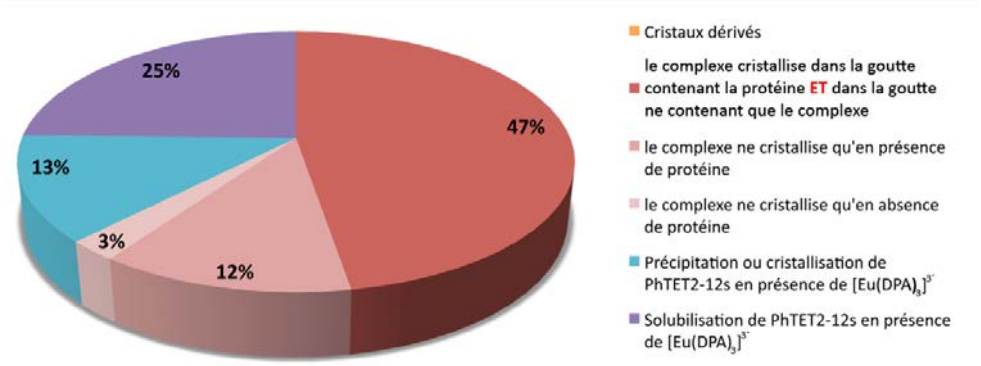

47%
12%
3%
13%
25%
Cristaux dérivés
le complexe cristallise dans la goutte contenant la protéine ET dans la goutte ne contenant que le complexe
le complexe ne cristallise qu'en présence de protéine
le complexe ne cristallise qu'en absence de protéine
Précipitation ou cristallisation de PhTET2-12s en présence de [Eu(DPA)3]3-
Solubilisation de PhTET2-12s en présence de [Eu(DPA)3]3-

Figure 3.12 – Bilan statistique des observations suite à l'ajout du complexe $[Eu(DPA)_3]^{3-}$ en co-cristallisation avec AfUox

En ce qui concerne la présence de cristaux dérivés et l'auto-cristallisation du complexe, le constat est le même qu'avec les protéines précédentes **[Figure 3.11]**. Dans 47 % des gouttes où le complexe $[Eu(DPA)_3]^{3-}$ est présent, il forme des cristaux de sel, qu'il soit en présence de protéine ou non. Dans 12 % des conditions de cristallisation, il ne cristallise que dans les gouttes qui contiennent la protéine, pas dans les gouttes équivalentes où il est seul, et ce, malgré le fait qu'il soit concentré à 100 mM. A l'inverse, dans 3 % des conditions de cristallisation, le complexe auto-cristallise dans les gouttes où il est seul mais pas dans les gouttes équivalentes avec la protéine **[Figure 3.12]**.

3.2.6 Co-cristallisation de la protéine PhTET2-12s avec le complexe $[Eu(DPA)_3]^{3-}$

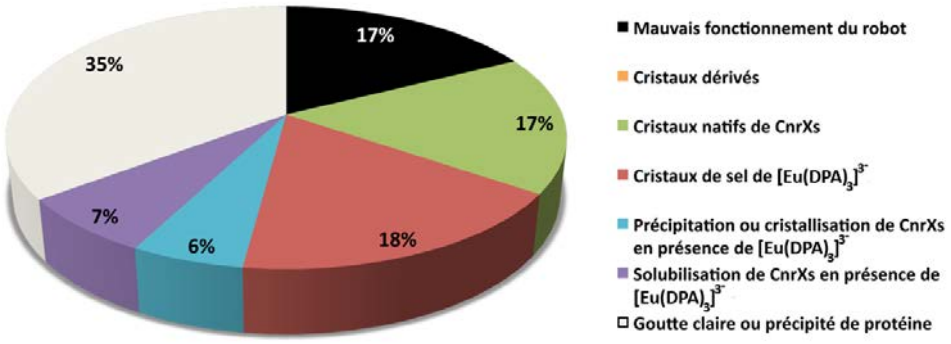


Figure 3.13 – Bilan statistique des observations des essais de co-cristallisation de PhTET2-12s avec le complexe $[Eu(DPA)_3]^{3-}$

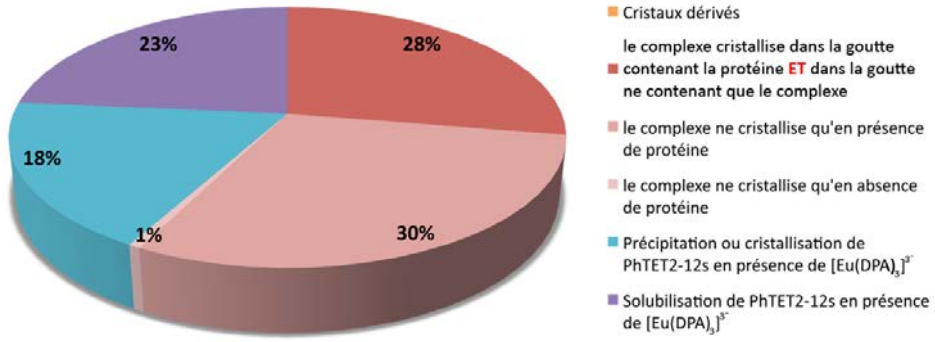

28%
30%
1%
18%
23%
Cristaux dérivés
le complexe cristallise dans la goutte contenant la protéine ET dans la goutte ne contenant que le complexe
le complexe ne cristallise qu'en présence de protéine
le complexe ne cristallise qu'en absence de protéine
Précipitation ou cristallisation de PhTET2-12s en présence de [Eu(DPA)3]3-
Solubilisation de PhTET2-12s en présence de [Eu(DPA)3]3-

Figure 3.14 – Bilan statistique des observations suite à l'ajout du complexe $[Eu(DPA)_3]^{3-}$ en co-cristallisation avec PhTET2-12s

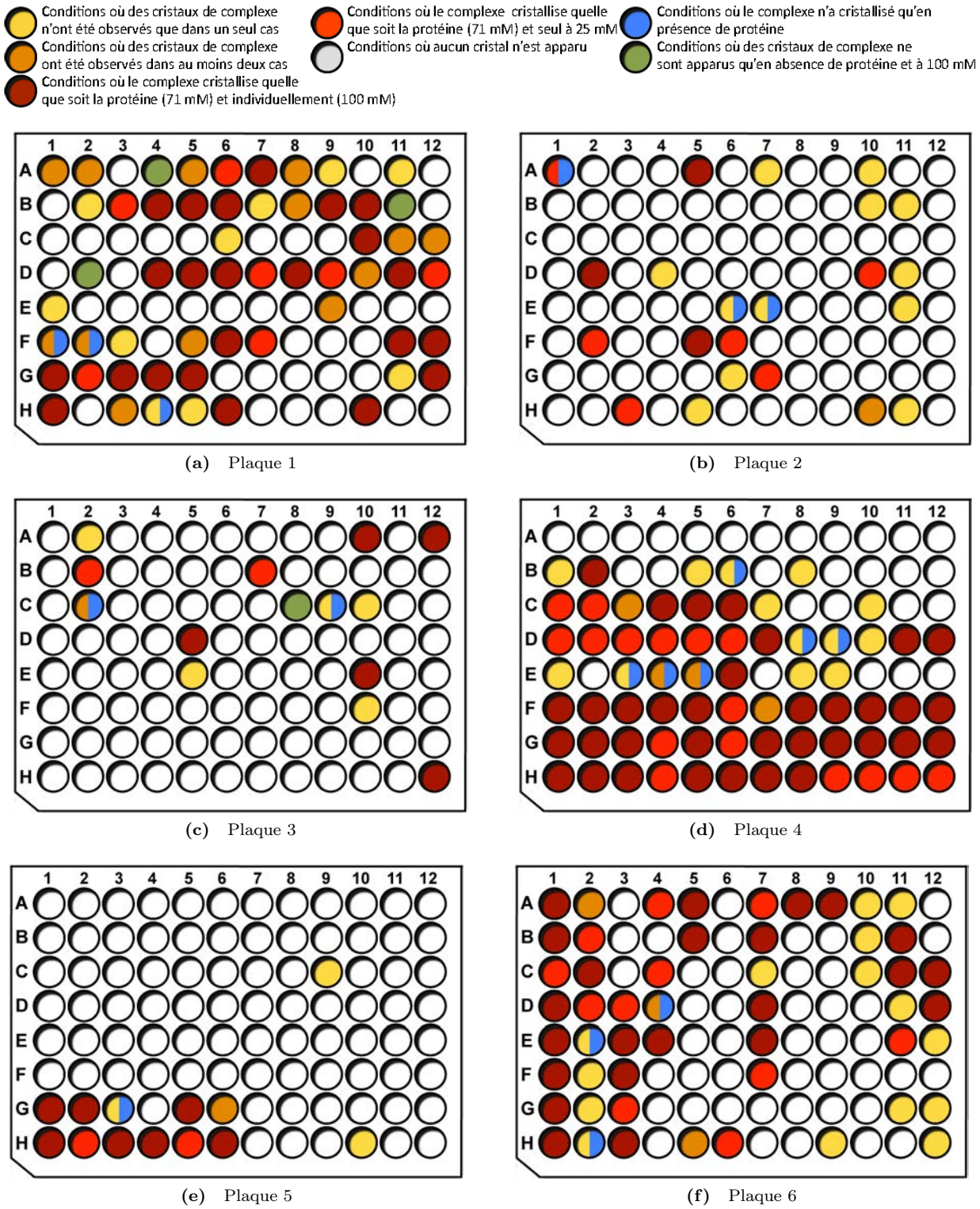
Cette expérience conduit aux même conclusions : la formation de cristaux de sel de complexe $[Eu(DPA)_3]^{3-}$ se produit dans 18% des conditions testées **[Figure 3.13]**, ce qui correspond toujours à la part la plus importante (58 %) des manifestations possibles du $[Eu(DPA)_3]^{3-}$ en co-cristallisation. Et concentré à 25 mM, seul dans la goutte, le complexe auto-cristallise beaucoup moins facilement qu'à 100 mM **[Figure 3.14]**.

3.3 Synthèse des observations et conclusion

En voulant vérifier de façon statistique, d'une part, l'hypothèse de l'utilisation du tris-dipicolinate de lanthanide en tant qu'agent facilitant la nucléation et, d'autre part, sa facilité d'emploi pour une expérience ordinaire de cristallisation automatisée compte tenu de sa propension à auto-cristalliser, j'ai testé le complexe en co-cristallisation avec six protéines dans 576 conditions de cristallisation, en goutte assise, à 20°C, à l'aide du robot nanogouttes HTX de l'EMBL. Aucun cristal dérivé n'a pu être observé. Ceci signifie que, lors de cette expérience, nous n'avons pas constaté d'action supra-moléculaire de la part

du tris-dipicolinate de lanthanide ou la manifestation caractéristique d'un agent capable de promouvoir la nucléation.

En regroupant toutes les observations faites sur les gouttes contenant du tris-dipicolinate d'europium avec les six protéines utilisées, on peut dresser un bilan statistique sur l'auto-cristallisation du complexe $[Eu(DPA)_3]^{3-}$. Ce bilan est illustré dans les **figure 3.15** et **3.16**.


Conditions où des cristaux de complexe n'ont été observés que dans un seul cas
Conditions où des cristaux de complexe ont été observés dans au moins deux cas
Conditions où le complexe cristallise quelle que soit la protéine (71 mM) et individuellement (100 mM)
Conditions où le complexe cristallise quelle que soit la protéine (71 mM) et seul à 25 mM
Conditions où aucun cristal n'est apparu
Conditions où le complexe n'a cristallisé qu'en présence de protéine
Conditions où des cristaux de complexe ne sont apparus qu'en absence de protéine et à 100 mM
1 2 3 4 5 6 7 8 9 10 11 12
A B C D E F G H

(a) Plaque 1 **(b)** Plaque 2

(c) Plaque 3 **(d)** Plaque 4

(e) Plaque 5 **(f)** Plaque 6

Figure 3.15 – Observations concernant l'auto-cristallisation du complexe $[Eu(DPA)_3]^{3-}$, toutes protéines confondues

Testé en co-cristallisation avec 6 protéines très différentes et à 20°C, le tris-dipicolinate d'europium a cristallisé au moins une fois dans 39 % des 576 conditions de cristallisation que composent les 6 kits de cristallisation commerciaux usuels. Il a cristallisé de façon reproductible, quelle que soit la protéine, dans 28 % de ces mêmes conditions **[Figure 3.16]**.

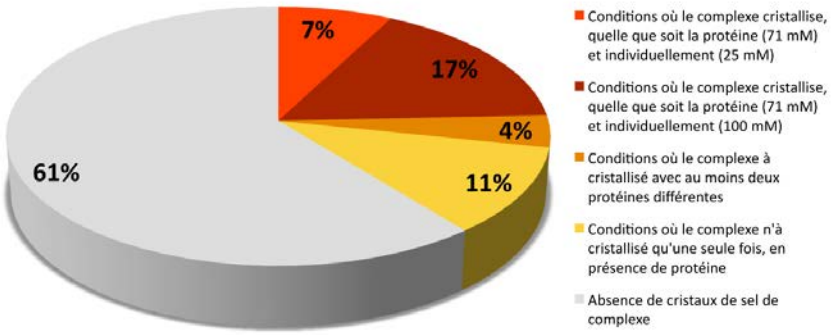

7%
17%
4%
11%
61%
Conditions où le complexe cristallise, quelle que soit la protéine (71 mM) et individuellement (25 mM)
Conditions où le complexe cristallise, quelle que soit la protéine (71 mM) et individuellement (100 mM)
Conditions où le complexe à cristallisé avec au moins deux protéines différentes
Conditions où le complexe n'à cristallisé qu'une seule fois, en présence de protéine
Absence de cristaux de sel de complexe

Figure 3.16 – Bilan statistique sur l'auto-cristallisation du complexe $[Eu(DPA)_3]^{3-}$ suite à des expériences de co-cristallisation menées avec 6 protéines différentes

La présence de protéine dans la goutte semble favoriser la formation de cristaux de sel de $[Eu(DPA)_3]^{3-}$, comme le montrent les expériences contrôles avec les protéines AfUox et PhTET2-12s **[Figure 3.16]**. Alors que nous recherchions l'effet supramoléculaire du $[Eu(DPA)_3]^{3-}$, nous avons parfois observé que sa présence dans la goutte provoquait l'effet inverse : un effet solubilisant. En effet, alors que des cristaux de protéine apparaissaient dans la goutte de cristallisation sans complexe, ils n'apparaissaient pas dans la goutte équivalente avec le complexe. Le complexe étant chargé trois fois négativement, il est possible qu'il puisse aussi provoquer le contraire de l'effet escompté : augmenter la solubilité de la protéine s'il s'y fixe. Une observation concernant non plus les cristaux mais les précipités corrobore cela. Les précipités de protéine en présence de $[Eu(DPA)_3]^{3-}$ sont systématiquement plus homogènes, moins diffus et presque dépourvus de volutes opaques par rapport aux précipités observables dans les gouttes avec la protéine seule.

Le nombre de conditions de cristallisation où apparaissent les cristaux de sel de complexe $[Eu(DPA)_3]^{3-}$ semble dépendant de la protéine testée. Bien que nous ayons exploré la piste du point isoélectrique de la protéine en solution, nous n'avons pas réussi à trouver de corrélation entre le taux d'auto-cristallisation du complexe et la protéine étudiée.

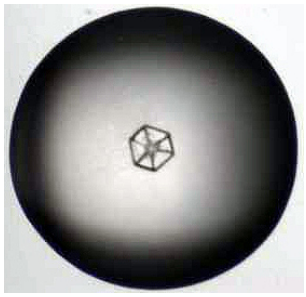
Nos observations confirment que le $[Eu(DPA)_3]^{3-}$ auto-cristallise systématiquement dans quatre situations :

- ▷ en présence d'ions divalents (Ca^{2+}, Mg^{2+}, Cd^{2+}, Ni^{2+}, *etc.*),
- ▷ dans des conditions où les concentrations en sels (chlorure, malonate et formate de sodium, sulfate ou acétate d'ammonium, chlorure de lithium, *etc.*) sont supérieures ou égales à 0,8 M,
- ▷ en présence du précipitant 2-Methyl-2,4-pentanediol, isopropanol,
- ▷ dès que la diffusion de vapeur est accentuée : en bordure de goutte ou dans une goutte qui s'assèche.

Les cristaux de sel de complexe sont reconnaissables car ils apparaissent rapidement, en quelques minutes à quelques heures et ils présentent une luminescence très intense une fois exposés aux UV proches. Les formes cristallines de sel de tris-dipicolinate de lanthanide diffèrent selon la protéine, le type de sel ou l'ion divalent présent dans la solution. Seul les cristaux en présence de calcium et en hautes concentrations en NaCl sont toujours reproductibles. La **figure 3.17** montre quelques exemples de cristaux de sel de $[Eu(DPA)_3]^{3-}$ caractéristiques, observés dans une goutte contenant de la protéine.

Les cristaux de sel de $[Eu(DPA)_3]^{3-}$ en présence de protéine sont d'ailleurs toujours différents de ceux qui apparaissent dans la même condition mais dépourvue de protéine : ils apparaissent toujours sous une forme plus "torturée", aux faces imparfaites, qui laisse moins bien passer la lumière. Ceci prouve qu'il y a une interaction en solution entre le complexe et les protéines.

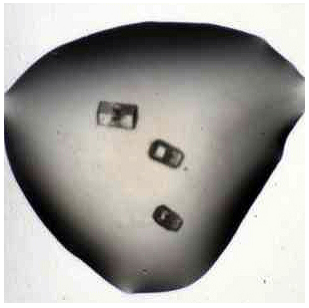
(a) Cristaux de sel de $[Eu(DPA)_3]^{3-}$ en présence de protéine et de Ca^{2+}



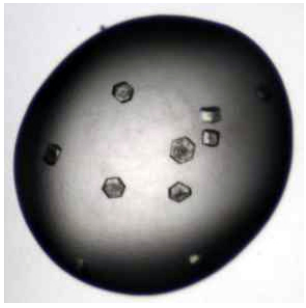
(b) Cristaux de sel de $[Eu(DPA)_3]^{3-}$ en présence de protéine et de $\geq$ NaCl 0,8 M

(c) Cristaux de sel de $[Eu(DPA)_3]^{3-}$ en présence de protéine et de Mg^{2+}

(d) Cristaux de sel de $[Eu(DPA)_3]^{3-}$ en présence de protéine et de $\geq$ citrate de sodium 0,8 M



(e) Cristaux de sel de $[Eu(DPA)_3m^{3-}$ en présence de protéine et de $\geq$ 0,8 M malonate de sodium 0,8 M



(f) Cristaux de sel de $[Eu(DPA)_3]^{3-}$ en présence de protéine et de $\geq$ 0,8 M malonate de sodium 0,8 M bis

Figure 3.17 – Formes cristallines caractéristiques de cristaux de sel de $[Eu(DPA)_3]^{3-}$

Pour conclure quant à l'utilisation de ce complexe pour obtenir des cristaux dérivés, je préconiserai la démarche suivante. Il faut en premier lieu vérifier si les conditions de cristallisation de la protéine native sont compatibles avec l'utilisation du complexe $[Ln(DPA)_3]^{3-}$. Si c'est le cas, commencer alors par utiliser ce complexe. Et si les conditions sont incompatibles avec son utilisation, lui préférer alors les complexes IRM.

Ces expériences de co-cristallisation automatisée ont donc montré qu'avec les six protéines différentes testées, nous n'avons jamais constaté l'effet supramoléculaire du complexe $[Ln(DPA)_3]^{3-}$. Pour essayer de déterminer quels sont les critères physico-chimiques et structuraux qui sont susceptibles de conduire à cet effet supramoléculaire, nous avons entrepris une nouvelle étude comparative sur deux structures résolues à l'aide du $[Ln(DPA)_3]^{3-}$ par Guillaume Pompidor au cours de ses travaux de recherche : le lysozyme de blanc d'œuf de poule ("HEWL") et le lysozyme de blanc d'œuf de dinde ("TEWL").

Dans le cas du HEWL, la fixation du complexe a entrainé le changement du groupe d'espace des cristaux dérivés apparus dans les conditions de cristallisation conduisant habituellement à la croissance de cristaux natifs. Ce phénomène ne s'est pas produit avec le TEWL qui pourtant cristallise dans les mêmes conditions que le HEWL et ne diffère de ce dernier que par la présence de sept substitutions. Afin de comprendre pourquoi dans le premier cas, il y a eu "effet supramoléculaire" et pas dans l'autre, nous avons entrepris cette nouvelle étude comparative. Les premières observations montrent que ce n'est pas l'acide aminé impliqué dans la liaison avec le $[Ln(DPA)_3]^{3-}$ qui est substitué mais celui avoisinant. Mener à terme cette étude comparative pourra peut être apporter des éclaircissements quant à l'effet supramoléculaire du tris-dipicolinate de lanthanide.

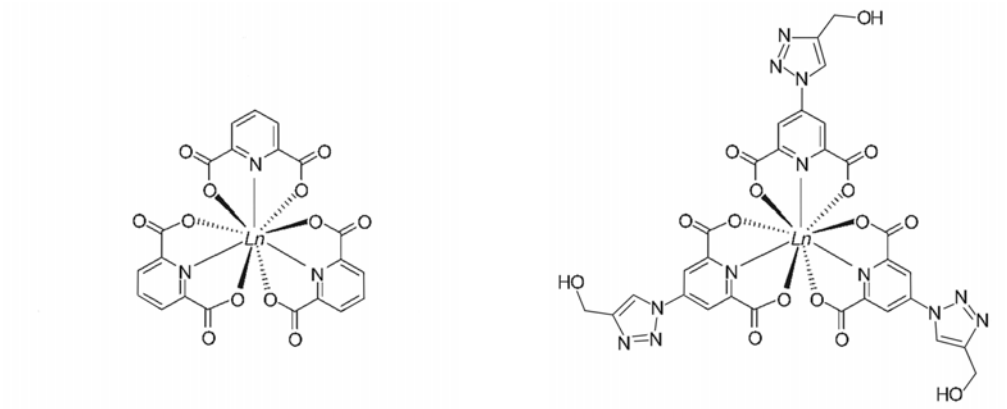
Chapitre 4

La famille des tris-triazoledipicolinates de lanthanide

Dans ce chapitre sont présentés le développement et l'étude de nouveaux complexes dérivés du complexe luminescent tris-dipicolinate de lanthanide qui a fait l'objet du chapitre précédent. Ces nouveaux complexes constituent la nouvelle famille des tris-triazoledipicolinates qui appartient donc à la grande famille des complexes picolinates de lanthanide.

4.1 Présentation du complexe luminescent $[Ln(HMTDPA)_3]^{3-}$

Le tris-hydroxyméthyltriazoledipicolinate de lanthanide est un complexe formé d'un cation lanthanide chélaté par trois anions 4-(4-(hydroxymethyl)-1H-1,2,3-triazol-1-yl)pyridine-2,6-dicarboxylate (aussi nommé "4-(4-hydroxymethyl-1H-V-triazole)dipicolinate" ou plus simplement "hydroxyméthyltriazoledipicolinate" : "HMTDPA"). Chacun des trois ligands établit une liaison dative avec l'atome de lanthanide à l'aide du doublet de l'atome d'azote de la pyridine et de celui de l'oxygène négativement chargé du groupement carboxylate. La sphère de coordination de l'atome de lanthanide est donc complète et la charge globale du complexe est triplement négative **[figure 4.1b]**.



(a) $[Ln(DPA)_3]^{3-}$ **(b)** $[Ln(HMTDPA)_3]^{3-}$

Figure 4.1 – Formule développée du complexe tris-dipicolinate de lanthanide (a) et du tris-hydroxyméthyltriazoledipicolinate de lanthanide (b)

Le $[Ln(HMTDPA)_3]^{3-}$ est une modification du $[Ln(DPA)_3]^{3-}$ dont la synthèse "par chimie-clic" (Lewis *et al.*, 2002; Chamas *et al.*, 2010) et les analyses spectroscopiques reviennent à Arnaud Goutier et Lionel Nauton, collaborateurs du laboratoire SEESIB à l'Université Blaise-Pascal de Clermont-Ferrand **[Figure 4.2]**.

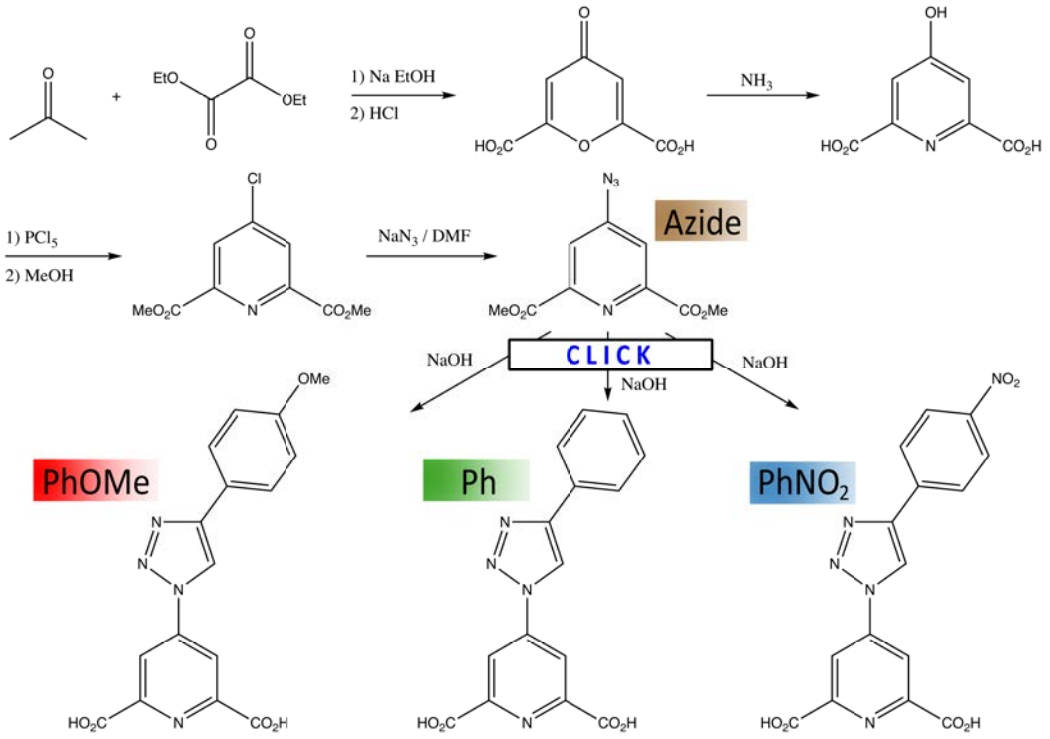


Figure 4.2 – Synthèse modulable d'une série de ligands de type triazoledipicolinate par chimie-clic

La figure 4.2 montre les trois premiers ligands ayant été obtenus au laboratoire SEESIB par chimie-click : "PhOMe" (acide 4-(4-(4-methoxyphenyl)-1H-1,2,3-triazol-1-yl)pyridine-2,6-dicarboxylique), "Ph" (acide 4-(4-phenyl-1H-1,2,3-triazol-1-yl)pyridine-2,6-dicarboxylique) et "$PhNO_2$" (acide 4-(4-(4-nitrophen yl)-1H-1,2,3-triazol-1-yl)pyridine-2,6-dicarboxylique). Les complexes de terbium et d'europium formés à l'aide de ces ligands ont fait l'objet d'analyses spectroscopiques qui seront décrites dans le prochain paragraphe.

4.1.1 Caractéristiques chimiques du complexe

Tout comme le complexe $[Ln(DPA)_3]^{3-}$, le $[Ln(HMTDPA)_3]^{3-}$ est une molécule non-cagée et chirale : on la retrouve donc sous forme d'un mélange racémique de deux formes énantiomères. La taille du complexe est équivalente à une sphère d'environ 22 Å de diamètre **[Figure 4.3]**. Sa solubilité dans l'eau MilliQ est inférieure à celle du tris-dipicolinate de lanthanide (220 mM à 298 K) : elle est de l'ordre de 75 mM à 298 K (température ambiante) et 110 mM à 315 K. Les complexes luminescents $[Eu(HMTDPA)_3]^{3-}$ et $[Tb(HMTDPA)_3]^{3-}$ nous ont été fournis sous forme de poudre avec trois ions sodium en guise de contre-cation. La solubilisation des poudres conduit à une solution transparente de couleur jaune. Plusieurs molécules d'eau, dont le nombre dépend du lanthanide chélaté, leur sont déjà associées. Le poids moléculaire d'un ligand $HMTDPA^{2-}$ est de 308,16 $g.mol^{-1}$.

Figure 4.3 – Les deux formes énantiomères du $[Ln(HMTDPA)_3]^{3-}$

Avant de disposer des complexes entiers, nous ne possédions que du ligand sous forme de poudre. J'ai pu alors préparer le complexe entier en suivant ce protocole :

Préparation du complexe $[Tb(HMTDPA)_3]^{3-}$ à partir du ligand seul

La masse molaire du ligand HMTDPA est égale à $M_{lig} = 308{,}16$ g.mol^{-1}. Après avoir pesé une masse $m_{lig} = 4{,}5$ mg de poudre, je l'ai dissoute entièrement dans un volume $V_{lig} = 95$ µL, ce qui conduit à une solution de concentration $C_{lig} = 153{,}7$ mM. Etant donné que le complexe est formé de trois ligands pour un ion lanthanide, il faudra une concentration finale de sel de lanthanide de $C_{Tb} = \frac{153{,}7}{3} = 50$ mM, un léger excès de ligand étant indispensable pour ne pas laisser d'ions lanthanides libres en solution. Ainsi, pour une masse $m_{Tb} = 1{,}4$ mg de chlorure de terbium pesée et pour $M_{Tb} = 373{,}38$ g.mol^{-1}, j'ai ajouté un volume $V_{lig} = \frac{m_{Tb}}{C_{Tb}M_{Tb}} = 75$ µL de solution de ligand pour préparer une solution contenant 50 mM de complexe $[Tb(HMTDPA)_3]^{3-}$. Dès lors, il est possible de vérifier la formation du complexe en solution qui fluoresce fortement dans le vert lorsqu'on l'expose aux proches UV ($\lambda = 315$ nm) à l'aide d'une lampe de paillasse (modèle qu'on utilise, par exemple, pour mettre en évidence l'ADN marqué au BET sur gel d'agarose) **[Figure 4.4]**.

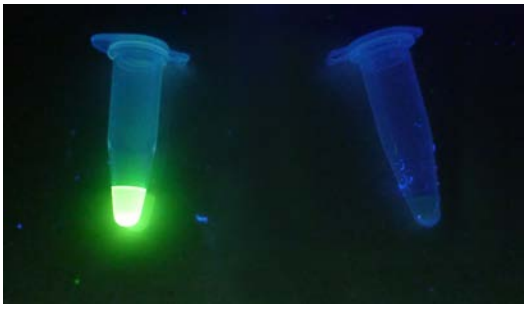


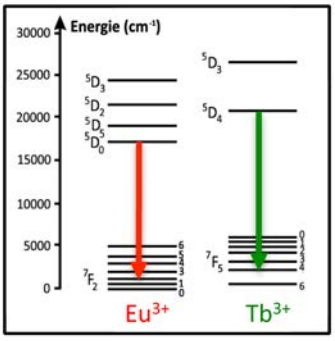
Figure 4.4 – Photo de tubes eppendorf® contenant une solution à 50 mM de complexe$[Tb(HMTDPA)_3]^{3-}$, à gauche, et d'une solution de ligand HMTDPA, à droite, sous exposition aux UV ($\lambda = 315$ nm) et en salle noire.

4.1.2 Caractéristique spectroscopique : la luminescence dans le visible par effet d'antenne

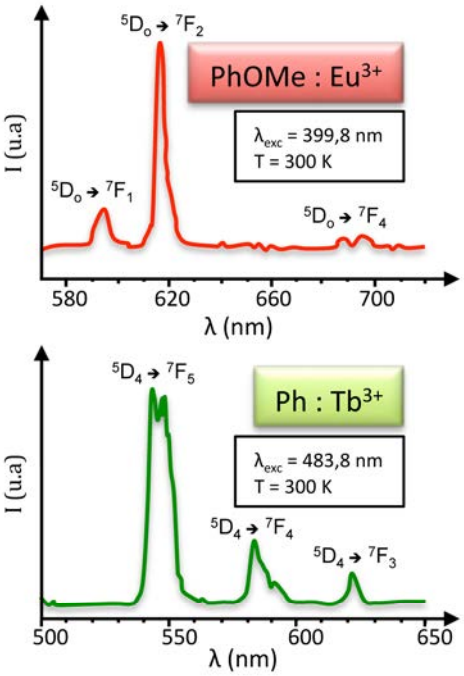
Ce sont les transitions des électrons de la couche f qui sont responsables de la luminescence naturelle des atomes de lanthanide (*cf.* introduction). En solution, le rendement de luminescence est très fortement atténué par la présence de molécules d'eau dans la première sphère de coordination du lanthanide. Le choix de travailler avec les cations Tb^{3+} et Eu^{3+} est du au fait que leurs raies d'absorption et d'émission sont situées respectivement dans le domaine de l'ultraviolet et du visible.

Comme pour le $[Tb(DPA)_3]^{3-}$ et le $[Eu(DPA)_3]^{3-}$, les trois ligands HMTDPA jouent le rôle de chromophores : ils absorbent la lumière et transfèrent une partie de l'énergie par effet d'antenne sur l'ion qui luminesce alors. La luminescence est d'autant plus visible que les ligands excluent les molécules d'eau de la sphère de coordination de l'ion, empêchant ainsi sa désexcitation non-radiative **[Figure 4.5]**.

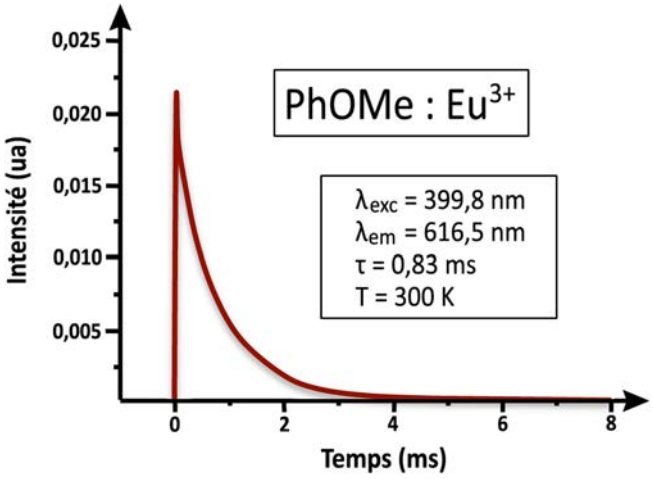
De plus, l'absorption du tris-dipicolinate de lanthanide dans le domaine UV est entièrement due à la fonction chimique pyridine-2,6-dicarboxylate. Celle du tris-hydroxyméthyltriazole-dipicolinate est non seulement due à cette même fonction mais le noyau triazole participe aussi à la délocalisation des électrons. La fluorescence dans le visible des complexes $[Ln(HMTDPA)_3]^{3-}$ est alors bien plus importante que celle du $[Ln(DPA)_3]^{3-}$ (nous l'estimons six à sept fois supérieure au regard de la fluorescence émise par des solutions de $[Tb(HMTDPA)_3]^{3-}$ et $[Tb(DPA)_3]^{3-}$ préparées à concentrations équivalentes).



(a) Diagramme d'énergie des états des électrons 4f des complexes $[Eu(HMTDPA)_3]^{3-}$ et $[Tb(HMTDPA)_3]^{3-}$



(b) Profil de fluorescence dans le visible des complexes $[Eu(HMTDPA)_3]^{3-}$ et $[Tb(HMTDPA)_3]^{3-}$



$PhNO_2$	0,85 ms
Ph	0,87 ms
N_3 (azide)	1,11 ms

(c) Profil des déclins de fluorescence

Figure 4.5 – Mesure de l'émission dans le visible des complexes de terbium et d'europium. Avec λ_{exc}, la longueur d'onde excitatrice, λ_{em}, la longueur d'onde d'émission, τ, le temps de demi-vie et T, la température

La **figure 4.6** montre qu'après adjonction des différentes fonctions chimiques par chimie-clic (à savoir les ligands "PhOMe", "$PhNO_2$" et "Ph", voir la **figure 4.2**), l'absorption dans l'UV-visible est nettement supérieure à celle constatée pour les complexes dont les ligands n'ont pas subi cette réaction (à savoir le ligand nommé "azide" (dimethyl 4-azidopyridine-2,6-dicarboxylate). En outre, cette expérience montre que la présence d'un groupement électro-attracteur (NO_2) provoque un déplacement du spectre d'absorption vers le domaine du visible.

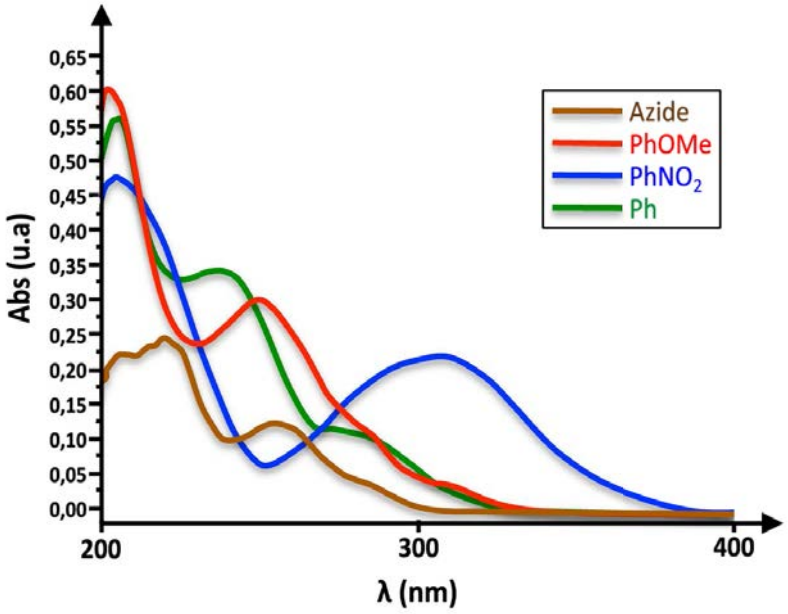


Figure 4.6 – Spectres d'absorption des complexes à différentes étapes de leur synthèse par spectroscopie dans l'UV-Visible

Comme nous le verrons par la suite, d'autres complexes similaires au $[Ln(HMTDPA)_3]^{3-}$ sont intervenus dans l'étude qui est présentée ici. L'ensemble de ces complexes forme la famille des tris-triazoledipicolinate de lanthanide ou "$[Ln(TDPA)_3]^x$" (où x est la charge des complexes qui varie d'un complexe à l'autre, *cf.* 4.6).

Nous avons entrepris une étude cristallographique systématique et comparative de cristaux dérivés de différentes protéines préparés avec plusieurs concentrations de complexe $[Ln(HMTDPA)_3]^{3-}$. Ceci pour, d'une part, déterminer si la résolution de la structure de ces protéines est possible par les méthodes tirant profit de la diffusion anomale et, d'autre part, déterminer quels sont alors les critères physico-chimiques requis pour la fixation du complexe.

Cette étude comprend d'abord l'obtention des cristaux natifs pour toutes les protéines, la préparation des cristaux dérivés puis l'acquisition et la réduction des données de diffraction, la détermination d'éventuels sites de fixation et le calcul des phases.

Dans un deuxième temps, pour les cristaux dérivés présentant des sites de fixation suffisamment occupés pour réaliser un phasage dit "expérimental", l'affinement du modèle structural de la protéine avec le complexe lié doit permettre d'identifier et de comprendre son mode de fixation. L'analyse des résultats de cette étude concernant la famille des $[Ln(TDPA)_3]^x$ et ses dérivés a pour but de comparer leurs modes d'interaction avec les protéines en fonction des données physico-chimiques et structurales dont on dispose : conditions de cristallisation, composition en acides aminés de la surface des protéines modèles, empilement cristallin, proportion de solvant, motifs structuraux particuliers. En les comparant aussi au $[Ln(DPA)_3]^{3-}$, nous devrions être en mesure de conclure quant aux avantages et aux inconvénients amenés par l'ajout des fonctions chimiques supplémentaires.

4.2 Préparation de cristaux dérivés avec le complexe $[Ln(HMTDPA)_3]^{3-}$ en vue d'une expérience de diffraction des rayons X

Une fois sa stabilité en solution vérifiée sur la durée, j'ai utilisé le $[Ln(HMTDPA)_3]^{3-}$ en co-cristallisation ou en trempage pour tenter d'obtenir des cristaux dérivés de six protéines différentes. Dans le cas de la co-cristallisation et dans le but d'expérimenter différents rapports de concentration, un volume de solution de complexe a été ajouté à des solutions concentrés de protéine ou directement sur des protéines lyophilisées pour les solubiliser.

Les concentrations de complexe $[Ln(HMTDPA)_3]^{3-}$ utilisées pour les trempages ont été de 2 mM à 27 mM. La durée de trempage a varié entre 2 minutes à plus d'1 heure, temps pendant lequel le cristal est observé sous binoculaire, afin de retirer le cristal de la solution en cas de détérioration visible. Toutes ces manipulations se font avec des boucles de taille adéquate.

4.2.1 Choix des protéines modèles et méthodes systématiquement utilisées

Cette étude systématique a été menée avec sept protéines différentes dont les conditions de cristallisation et la structure ont été respectivement déterminées. Les protéines choisies ont été, pour la plupart, les mêmes que celles que nous avons utilisées pour l'étude des propriétés de nucléation du $[Ln(DPA)_3]^{3-}$ présentée dans le chapitre précédent.

Outre celles déjà fournies par nos collaborateurs, je rappelle ici les critères de choix d'une protéine modèle pour tester l'utilisation des complexes. La protéine ayant un rôle de modèle sur lequel on applique plusieurs méthodes à étudier, il est nécessaire de disposer d'une quantité suffisante de protéine pure et de connaitre les conditions de cristallisation qui permettent d'obtenir des cristaux le plus rapidement possible qui diffractent à relativement haute résolution. En effet, la haute résolution est indispensable pour une description précise des interactions entre les complexes et les protéines. Un autre critère est de choisir un ensemble de protéines représentant un éventail éclectique de tailles, de propriétés de surface (charges, cavités...) et de propriétés cristallines (proportion de solvant, empilement). Il convient d'ajouter à ces critères la possibilité de disposer de protéines capable de cristalliser dans différentes conditions, ce qui doit permettre de déterminer l'influence des conditions physico-chimiques de cristallisation (température, pH, agent précipitant) sur la fixation des complexes. Enfin, un autre critère peut être de préférer des protéines cristallisant dans un groupe d'espace de haute symétrie afin de bénéficier d'une redondance élevée et ainsi améliorer le rapport signal sur bruit conduisant à une estimation des différences anomales encore plus précise.

Dans notre cas, bien que plusieurs conditions de cristallisation aient été décrites pour chacune des protéines choisies, nous n'avons toujours pu les cristalliser que dans une seule de leurs conditions décrites.

4.2.2 Le lysozyme de blanc d'œuf de poule ("HEWL" pour Hen Egg White Lysozyme)

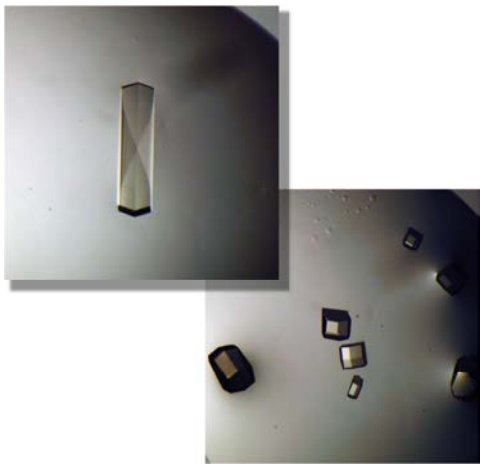
Le lysozyme ("Murramidase" ou encore " N-acetylmuramide glycanhydrolase") est une enzyme globulaire soluble eucaryote qui appartient à la famille des O-glycosyl hydrolases (classification enzymatique E.C. 3.2.1.17). Son poids moléculaire est de 14,3 kDa pour 129 acides aminés, avec un point isoélectrique

théorique de 9,32 et un coefficient d'extinction molaire théorique à 280 nm $\varepsilon_{280\,nm} = 37970\ M^{-1}.cm^{-1}$ (Grunclová *et al.*, 2003).

Nous utilisons la protéine commercialisée par la société Roche® sans purification ni dialyse supplémentaire.

4.2.2.1 Cristallisation de la protéine native

Le lysozyme de blanc d'œuf de poule cristallise dans un certain nombre de conditions physico-chimiques qui conduisent à des formes cristallines différentes. Je me suis inspiré des conditions habituellement utilisées pour obtenir des cristaux de forme quadratique le plus rapidement possible (Ducruix & Giegé, 1992) **[Figure 4.7]**.



(a) Cristaux natifs de HEWL, ≈ 1330x200 µm - 380x240 µm

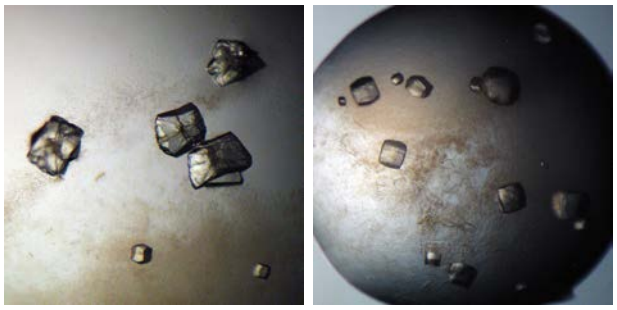
0,4 - 1,5 M NaCl
100 mM acétate de sodium pH 4,6
[P] = 20 mg.mL
Goutte : 1,5 µL (P) + 1,5 µL (R)
T = 293 K (20°C)
3 à 7 jours (12h si [P] =100 $mg.mL^{-1}$)
Cryo-protection : (R) + 25 % glycérol +/- parafine

(b) Conditions de cristallisation et de cryo-protection

Figure 4.7 – Cristallisation et cryo-protection du lysozyme de blanc d'œuf de poule

4.2.2.2 Obtention des cristaux dérivés

Des cristaux dérivés de HEWL ont pu être obtenus par trempage et par co-cristallisation avec le $[Ln(HMTDPA)_3]^{3-}$ dans les mêmes conditions de cristallisation que la protéine native mais dans la partie haute de la gamme de concentration en NaCl **[Figure 4.8]**. En co-cristallisation, comme observé pour les cristaux dérivés de HEWL avec le complexe $[Ln(DPA)_3]^{3-}$ (Pompidor *et al.*, 2010), la nucléation se produit à partir d'un précipité blanchâtre et dense qui se forme dès l'ajout du complexe dans la goutte. Ce dernier disparait progressivement au fur et à mesure que les cristaux croissent.



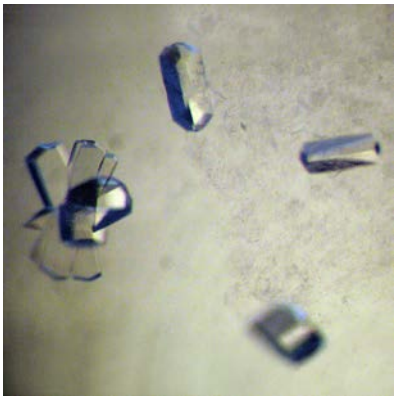
(a) Cristaux dérivés de HEWL, ≈ 240x200 µm c 100x90 µm

0,9 - 1,5 M NaCl
100 mM acétate de sodium pH 4,6
$[P]_{stock}$ = 20 mg.mL^{-1}
$[C]_{stock}$ = 30 - 75 mM
Goutte : 1,5 µL (P) + 1,5 µL (R) + 1,5 µL (C) ou
2 µL (P) + 2 µL (R) + 1 µL (C) ou
2 µL (P) + 5 µL (R) + 2 µL (eau MQ) + 1 µL (C)
T = 293 K (20°C)
5 (1,3 - 1,5 M NaCl) à 15 jours (0,9 M NaCl)
Cryo-protection : (R) + 25 % glycérol +/- parafine

(b) Conditions de cristallisation et de cryo-protection

Figure 4.8 – Les deux formes cristallines des cristaux dérivés HEWL + $[Ln(HMTDPA)_3]^{3-}$ en co-cristallisation

La fixation du complexe dans les cristaux est aisément vérifiable sous binoculaire à l'aide d'une lampe à UV de paillasse en exposant la goutte aux UV proches (λ = 315 nm). Le complexe étant plus fortement concentré dans les cristaux que dans la solution de la goutte proprement-dite, le contraste entre la faible luminescence émise par la solution et celle des cristaux, bien plus intense dans le visible, permet de déterminer si le complexe s'est fixé ou non **[Figure 4.9]**.



(a) 3 mg.mL^{-1} HEWL + 2,7 mM $[Tb(HMTDPA)_3]^{3-}$, lumière visible

(b) 3 mg.mL^{-1} HEWL + 2,7 mM $[Tb(HMTDPA)_3]^{3-}$, proches UV

Figure 4.9 – détection du $[Tb(HMTDPA)_3]^{3-}$ dans des cristaux dérivés de HEWL

Pour diminuer la densité du précipité dans la goutte qui se forme dès l'ajout du complexe, j'ai parfois eu recours à une modification des proportions de protéine, de complexe et de réservoir dans la goutte de cristallisation. J'ai ajouté pour cela différents volumes de solution de protéine, de complexe et de réservoir dans la goutte, modifiant ainsi leur dilution respective **[figure 4.8]**. Afin d'éviter toute confusion, le rapport "quantité de complexe sur quantité de protéine à t_0 dans la goutte" sera indiqué pour chaque tableau de données de ce chapitre.

En co-cristallisation, on constate la présence de deux formes cristallines dans la goutte **[Figure 4.8]** qui luminescent sous exposition aux UV. La première forme ressemble à des cristaux natifs mais "torturés" et aux faces striées. Elle apparaît la première et la taille des cristaux est généralement assez importante. La deuxième forme ressemble en tout point aux cristaux natifs de HEWL. Ils apparaissent quelques jours après les précédents et sont de taille réduite.

Que ce soit en trempage ou en co-cristallisaton, il m'a été possible d'obtenir des cristaux dérivés de lysozyme de blanc d'œuf de poule avec différentes concentrations de complexe. Bien que le complexe $[Ln(HMTDPA)_3]^{3-}$ait été utilisé à de faibles concentrations, sa luminescence intrinsèque dans le visible étant supérieure à celle du $[Ln(DPA)_3]^{3-}$, il a été relativement facile de détecter sa fixation dans les cristaux en les exposant aux UV proches. Lorsqu'aucun signal anomal n'a pu être détecté suite à l'enregistrement de données cristallographiques sur l'un de ces cristaux, la luminescence des autres cristaux du même type et de la même goutte que celui utilisé pour l'expérience de diffraction fut vérifiée.

En outre, bien que deux faciès cristallins soient observables dans le cas de cristaux dérivés obtenus en co-cristallisation, celui "torturé" semble conserver la forme générale d'un cristal natif : je n'ai jamais observé avec le $[Ln(HMTDPA)_3]^{3-}$ de changement radical de forme cristalline comme cela a pu être observé avec le $[Ln(DPA)_3]^{3-}$ (*cf.* introduction).

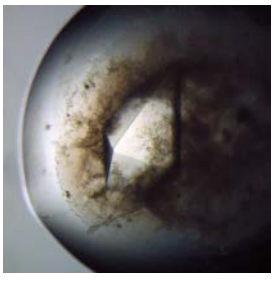
4.2.3 La thaumatine de *Thaumatococcus daniellii* "TdTHAU"

La thaumatine est une protéine globulaire soluble eucaryote qui est extraite du cytoplasme des cellules qui constituent les graines du fruit de l'arbre *Thaumatococcus daniellii*. Composée de 207 acides aminés, son poids moléculaire est de 22,2 kDa, pour un point isoélectrique théorique de 8,5 et un coefficient d'extinction molaire à 280 nm $\varepsilon_{280\,nm} = 29420\ M^{-1}.cm^{-1}$.

J'ai utilisé la protéine commercialisée par la société Sigma-Aldrich®, (référence T7638), sans purification ni dialyse supplémentaires.

4.2.3.1 Cristallisation de la protéine native

Tout comme le lysozyme, il est possible d'obtenir des cristaux de différents groupes d'espace dans différentes conditions (Charron *et al.*, 2004). Je n'ai réussi à obtenir des cristaux que dans des conditions inspirées de celles décrites par Ko *et al.* (1994). La nucléation se produit toujours à partir d'un léger précipité [**Figure 4.10**].



(a) Cristal natif de TdTHAU, 1100x520 µm

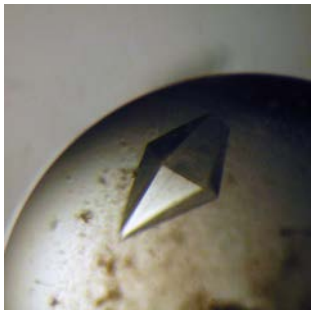
0,3 - 0,9 M tartrate double de potassium et de sodium
100 mM bis-tris propane pH 6,5
[P] = 40 $mg.mL^{-1}$
Goutte : 1,5 µL (P) + 1,5 µL (R)
T = 293 K (20°C)
2 à 7 jours
Cryo-protection : 25 % éthylène glycol ou paratone

(b) Conditions de cristallisation et de cryo-protection

Figure 4.10 – Cristallisation et cryo-protection de la thaumatine

4.2.3.2 Obtention des cristaux dérivés

Des cristaux dérivés de TdTHAU ont pu être obtenus par co-cristallisation avec le $[Ln(HMTDPA)_3]^{3-}$ et ce, dans les mêmes conditions de cristallisation que la protéine native. Les cristaux semblent identiques en tout point aux cristaux natifs **[Figure 4.11].**



(a) Cristaux dérivés de TdTHAU, ≈ 700x190 µm

0,3 - 0,9 M tartrate double de potassium et de sodium
100 mM bis-tris propane pH 6,5
$[P]_{stock}$ = 40 mg.mL^{-1}
$[C]_{stock}$ = 30 - 75 mM
Goutte : 1,5 µL (P) + 1,5 µL (R) + 1,5 µL (C)
T = 293 K (20°C)
5 jours (0,9 M tartrate K/Na) à 10 jours (0,3 M)

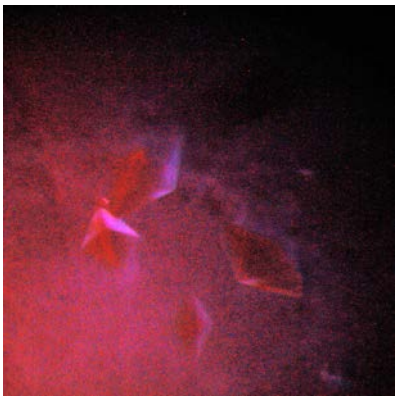
Cryo-protection : 25 % éthylène glycol ou paratone

(b) Conditions de cristallisation et de cryo-protection

Figure 4.11 – Cristaux dérivés de TdTHAU + $[Ln(HMTDPA)_3]^{3-}$ par co-cristallisation

En comparasaison à une goutte de cristallisation sans complexe et quelle que soit la concentration de précipitant, j'ai systématiquement dénombré davantage de cristaux de taille plus réduite dans une goutte de cristallisation avec complexe. La présence de ce dernier semble augmenter le nombre de points de nucléation. La fixation du complexe dans les cristaux est moins facilement vérifiable lors d'une exposition aux UV, car non seulement les cristaux luminescent, mais le précipité qui les entoure aussi. De fait, le contraste entre la luminescence émise par le précipité et celle des cristaux est un peu moins discernable sous binoculaire **[Figure 4.12]**.

(a) 20 mg.mL^{-1} TdTHAU + 4.5 mM $[Eu(HMTDPA)_3]^{3-}$, lumière visible



(b) 20 mg.mL^{-1} TdTHAU + 4.5 mM $[Eu(HMTDPA)_3]^{3-}$, proches UV

Figure 4.12 – Détection du $[EuHMTDPA)_3]^{3-}$ dans des cristaux dérivés de TdTHAU

Il m'a été possible d'obtenir des cristaux dérivés de thaumatine en co-cristallisation ou en trempage avec différentes concentrations de complexe. Seuls les cristaux dérivés obtenus par co-cristallisation ont permis un phasage expérimental.

4.2.4 La Protéine Fluorescente Verte de *Aequorea victoria* ("AvEGFP" pour *Aequorea victoria* Enhanced Green Fluorescent Protein)

Nous avons une nouvelle fois choisi la AvEGFP, protéine qui a déjà été utilisée lors de l'étude consignée dans le chapitre précédent. La protéine a été produite et purifiée au laboratoire à partir du clone bactérien transformé que notre collaborateur M. Philippe Oger, du Laboratoire de Sciences de la Terre à l'Ecole Normale Supérieure de Lyon, nous avait confié. Sa description ainsi que son protocole de production et de purification seront donnés en annexe.

4.2.4.1 Cristallisation de la protéine native

Malgré tout le soin apporté aux tentatives de reproduction des expériences de cristallisation décrites dans la litterature (Ormö *et al.*, 1996; Elsliger *et al.*, 1999; Rosenow *et al.*, 2004) et malgré les vérifications idoines quant à l'intégrité de la protéine (*cf.* annexes), aucune d'elles n'a conduit à une cristallisation reproductible de la protéine. C'est en menant l'expérience de co-cristallisation automatisée à l'aide du robot de cristallisation nanogoutte de l'EMBL, décrite dans le chapitre précédent, que nous avons pu déterminer une nouvelle condition de cristallisation pour l'AvEGFP [**Figure 4.13**].

(a) Cristaux natifs d'EGFP, 575x250 µm

70 -75 % MPD
100 mM HEPES pH 7,9
[P] = 20 $mg.mL^{-1}$
Goutte : 2 µL (P) + 2 µL (R)
T = 293 K (20°C)
7 à 10 jours
Cryo-protection : la liqueur mère est déjà cryo-protectante

(b) Conditions de cristallisation et de cryo-protection

Figure 4.13 – Cristallisation et cryo-protection de l'EGFP

4.2.4.2 Obtention de cristaux dérivés

Les essais de co-cristallisation et de trempage avec le $[Ln(HMTDPA)_3]^{3-}$ se sont tous révélés infructueux. Dès l'ajout du complexe dans une telle concentration en MPD, un précipité cristallin se forme. Ainsi, tout comme le tris-dipicolinate de lanthanide, le $[Ln(HMTDPA)_3]^{3-}$ est inutilisable en présence de MPD.

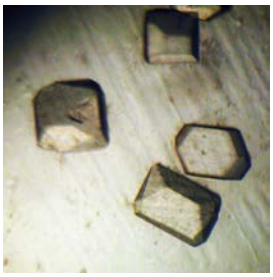
4.2.5 L'urate oxydase de *Aspergillus flavus* "AfUox"

Tout comme le fut la protéine AvEGFP, nous réutilisons la protéine AfUox pour cette étude. L'urate oxydase est une enzyme homotétramèrique de 137 kDa (4x34,24 kDa) composée de 1204 acides aminés (4x301). Son point isoélectrique théorique est de 7,18 et son coefficient d'extinction molaire est de $\varepsilon_{280\,nm} = 53525\ \mathrm{M^{-1}.cm^{-1}}$.

Un aliquot de 400 µL de solution comprenant la protéine concentrée à 11,3 $\mathrm{mg.mL^{-1}}$, un inhibiteur (8-azaxanthine) à 0,28 $\mathrm{mg.mL^{-1}}$ ainsi qu'une plaque contenant déjà des cristaux, obtenus en présence de l'inhibiteur et selon la méthode "batch" (*cf.* introduction), m'ont été très généreusement confiés par le Dr. Nathalie Colloc'h de l'équipe CERVOxy du laboratoire Imagerie et Stratégie Thérapeutiques des pathologies Cérébrales et Tumorales, à Caen et par le Pr. Thierry Prangé du Laboratoire de Cristallographie et RMN Biologiques de l'Université Paris V René Descartes, à Paris.

4.2.5.1 Cristallisation de la protéine native

Les conditions de cristallisation utilisées sont inspirées de celles décrites par Vivarès et Bonneté (Bonnete *et al.*, 2001) [**Figure** 4.14].



(a) Cristaux natifs d'AfUox, 480x290 µm

5 - 15 % PEG 8000
100 mM Tris HCl pH 8,5
[P] = 11,3 $\mathrm{mg.mL^{-1}}$
Goutte : 5 µL (P) + 5 µL (R)
T = 293 K (20°C)
7 à 8 jours
Cryo-protection : 25 % PEG 400

(b) Conditions de cristallisation et de cryo-protection

Figure 4.14 – Cristallisation et cryo-protection de l'AfUox

4.2.5.2 Obtention des cristaux dérivés

Les essais de co-cristallisation avec le $[\mathrm{Ln(HMTDPA)_3}]^{3-}$ se sont tous révélés infructueux car la présence du complexe conduit à une absence de cristallisation. Seule la méthode de trempage, d'une durée inférieure ou égale à 20 minutes et avec une concentration en complexe inférieure ou égale à 10 mM, a été possible. En effet, pour une durée de trempage supérieure à 20 minutes et/ou pour une concentration en complexe supérieure à 15 mM, les cristaux se dissolvaient.

4.2.5.3 Recherche de conditions de co-cristallisation à l'aide du robot de cristallisation nanogoutte de l'EMBL

Devant l'impossibilité d'obtenir des cristaux dérivés par co-cristallisation au laboratoire et profitant de l'expérience de co-cristallisation automatisée décrite dans le chapitre précédent, j'ai recherché d'autres conditions de co-cristallisation potentielles pour l'échantillon de AfUox.

Trois échantillons ont donc été menés au robot de cristallisation nanogoutte de l'EMBL :

1. 90 µL de solution comprenant 7,5 $mg.mL^{-1}$ de AfUox + 0,18 $mg.mL^{-1}$ de 8-azaxanthine,
2. 90 µL de solution comprenant 7,5 $mg.mL^{-1}$ de AfUox + 0,18 $mg.mL^{-1}$ de 8-azaxanthine + 10 mM $[Eu(HMTDPA)_3]^{3-}$,
3. 90 µL de solution comprenant 20 mM $[Eu(HMTDPA)_3]^{3-}$.

Comme je l'ai mentionné dans le premier chapitre de cette première partie, le robot utilise les plaques de cristallisation Greiner® de 96 puits avec 3 emplacements pour gouttes assises par puits. J'ai donc pu, en une seule expérience :

- ▷ rechercher des conditions de co-cristallisation pour la protéine avec le $[Eu(HMTDPA)_3]^{3-}$,
- ▷ rechercher de nouvelles conditions de cristallisation pour la solution de AfUox utilisée,,
- ▷ étudier, comme cela avait été fait pour le $[Ln(DPA)_3]^{3-}$, l'influence des 576 conditions de cristallisation des kits commerciaux sur le complexe, en présence ou non de protéine.

Les kits commerciaux qui ont été utilisés pour cette expérience sont : "The Classics (Qiagen® / Nextal®)" pour la plaque 1, "Crystal Screen light (Hampton®)"pour la plaque 2, "MemFac (Hampton®)" pour la plaque 3, "Crystal Ammonium Sulfate (Hampton®)" pour la plaque 4, "PEG 6K (Hampton®)" pour la plaque 5 et "Screen Index (Hampton®)" pour la plaque 6.

Mener cette expérience ne m'a pas permis de mettre en évidence des conditions de cristallisation ou de co-cristallisation réellement différentes de celle mentionnée ci-dessus : les cristaux les plus prometteurs apparaissant dans des conditions très similaires à celle-ci.

Toutefois, à travers cette expérience, j'ai pu confirmer que, tout comme le $[Ln(DPA)_3]^{3-}$, le complexe $[Ln(HMTDPA)_3]^{3-}$ n'est pas utilisable en présence de fortes concentrations en sel ($\geq$ 1 M), d'ions divalents (Ca^{2+}, Mg^{2+}, Li^{2+}...) ou de MPD ($\geq$ 25 %) car il cristallise ou précipite dans ces conditions. En outre, les cristaux de sel de complexe $[Ln(HMTDPA)_3]^{3-}$ observés diffèrent de ceux de sel de $[Ln(DPA)_3]^{3-}$. Le complexe $[Ln(HMTDPA)_3]^{3-}$ semble former plus facilement un précipité cristallin de sel de complexe que de véritables cristaux.

4.2.6 La malate déshydrogénase de *Chloroflexus aurantiacus* "CaMalDH"

Pour cette étude cristallographique des complexes, $[Ln(TDPA)_3]^x$, j'ai voulu utiliser la malate déshydrogénase de *Chloroflexus aurantiacus* que nous avons utilisée pour une autre étude décrite dans le second chapitre de la seconde partie de ce manuscrit. Cette enzyme m'a été fournie purifiée et concentrée à 20 $mg.mL^{-1}$ par le Dr. Dominique Madern, membre de notre groupe E.L.M.A..

4.2.6.1 Cristallisation de la protéine native

Les conditions de cristallisation utilisées sont inspirées de celles décrites par Dalhus *et al.* (2002) et des observations issues de l'expérience de co-cristallisation automatisée décrite dans le précédent chapitre. Toutes les informations relatives à la protéine et à sa cristallisation se trouvent dans le second chapitre de la seconde partie de ce manuscrit.

4.2.6.2 Obtention des cristaux dérivés

Les essais de co-cristallisation ou de trempage avec le complexe se sont tous révélés infructueux. La présence de 40 mM d'acétate de cadmium dans les conditions de cristallisation de la protéine provoque l'apparition immédiate d'un précipité cristallin opaque de sel de complexe.

Dans un premier temps, j'ai voulu essayer d'obtenir des cristaux dérivés dans d'autres conditions de cristallisation. Mais rares sont celles où la protéine semble cristalliser sans qu'il n'y ait de sel d'ions divalents. Dans ces rares conditions, les profils de diffraction des cristaux obtenus n'ont pas permis d'enregistrer des données.

Nous avons aussi essayé de co-cristalliser la protéine avec une concentration suffisamment faible de complexe dans l'espoir que la totalité des molécules de complexe se fixe aux molécules de protéine avant l'étape de nucléation empêchant ainsi le complexe de précipiter. J'ai donc réalisé des essais de co-cristallisation avec les deux rapports suivants : 1 molécule de CaMalDH : 1,5 molécules de $[Ln(HMTDPA)_3]^{3-}$ et 1 molécule de CaMalDH : 10 molécules de $[Ln(HMTDPA)_3]^{3-}$. Ceci m'a permis d'obtenir des cristaux semblables aux cristaux natifs. Mais en observant les mêmes gouttes plusieurs semaines après, j'ai constaté l'apparition d'un précipité cristallin opaque isolé, identique à celui observé dans les gouttes de trempage. Exposés aux UV proches, ces cristaux ne se sont pas révélés luminescents, confirmant davantage encore que le complexe ne s'était pas fixé à la protéine. D'après le calcul permettant de déterminer la quantité de complexe introduit par rapport à la quantité de protéine utilisée (*cf.* introduction), calcul réalisé après-coup, nous étions peut-être toujours en excès de complexe alors que nous pensions être en proportions stœchiométriques 1 :1.

4.2.7 La malate déshydrogénase de *Haloferax volcanii* "HvMalDH"

Comme pour l'enzyme CaMalDH, nous avons voulu profiter du fait que, dans le cadre de l'étude sur les malate déshydrogénases, nous disposions de la malate déshydrogénase de *Haloferax volcanii.* Cette enzyme a donc été naturellement choisie comme protéine modèle pour l'étude de la famille des $[Ln(TDPA)_3]^x$. Comme CaMalDH, la protéine m'a été fournie purifiée et concentrée à 10 mg.mL^{-1} par le Dr. Dominique Madern.

4.2.7.1 Cristallisation de la protéine native

Les conditions de cristallisation de la protéine HvMalDH étant inconnues, nous avons eu recours, une fois de plus, à une expérience de co-cristallisation automatisée menée à l'aide du robot de cristallisation nanogoutte de l'EMBL. Tout comme la description biochimique de la protéine, les conditions de cristallisation de la protéine native sont décrites dans le second chapitre de la deuxième partie de ce manuscrit.

4.2.7.2 Obtention des cristaux dérivés

Ne disposant que d'une petite quantité de protéine, seules des expériences de trempage ont pu être réalisées avec la HvMalDH et le complexe $[Ln(HMTDPA)_3]^{3-}$. D'une durée inférieure ou égale à 5 min avec une solution contenant 20 mM $[Ln(HMTDPA)_3]^{3-}$ dépourvue de $MgCl_2$ (présent dans les conditions de cristallisation de la protéine native), ces trempages ont pu être menés à bien.

4.2.8 L'aminopeptidase TET1-12s de *Pyrococcus horikoshii* "PhTET1-12s"

L'aminopeptidase TET1-12s de *Pyrococcus horikoshii* a aussi été choisie comme protéine modèle pour cette étude. Une fois de plus, nous disposions déjà de la protéine pour la co-cristalliser en présence du complexe $[Eu(DPA)_3]^{3-}$ en vue de l'expérience de diffraction de rayons X décrite dans le premier chapitre de la seconde partie. La protéine m'a été fournie purifiée et concentrée à 6,2 $mg.mL^{-1}$ par la Dr. Asunción Durá, membre de l'équipe (Durá *et al.*, 2005).

4.2.8.1 Cristallisation de la protéine native

Les conditions de cristallisation sont celles décrites dans le premier chapitre de la seconde partie de ce manuscrit. Elles m'ont été transmises par la Dr. Asunción Durá.

4.2.8.2 Obtention des cristaux dérivés

Ne disposant que d'une petite quantité de solution concentrée en protéines, seuls des essais par trempage ont pu être réalisés avec la PhTET1-12s. J'ai réalisé des trempages de 2 à 5 minutes avec une solution comprenant 10 mM $[Ln(HMTDPA)_3]^{3-}$. Les cristaux de PhTET112s n'ont pas été détériorés par cette méthode.

4.2.9 Constats expérimentaux concernant la préparation de cristaux dérivés avec le $[Ln(HMTDPA)_3]^{3-}$

Bien qu'il fut toujours possible d'obtenir des cristaux natifs de protéine, les essais de cristallisation n'ont pas tous permis d'obtenir des cristaux dérivés exploitables. Les tentatives d'obtention de cristaux dérivés avec les protéines CaMalDH et AvEGFP ont montré que, comme le tris-dipicolinate de lanthanide dont il est issu, le tris-hydroxyméthyltriazole-dipicolinate de lanthanide est inutilisable en présence de Cd^{2+} ou en présence de 2-methyl-2,4-pentanediol. Utiliser le $[Ln(HMTDPA)_3]^{3-}$ dans ces conditions provoque l'apparition d'un précipité cristallin opaque et fortement luminescent. Suite à l'expérience de co-cristallisation automatisée réalisée avec l'urate oxidase de *Aspergillus flavus*, nous avons pu confirmer cette observation et l'étendre aux conditions de cristallisation composées de hautes concentrations salines, d'ions Mg^{2+}, ou d'ions Ca^{2+}.

Dérivé chimique du $[Ln(DPA)_3]^{3-}$, on pouvait s'attendre à un tel résultat mais à la différence de ce dernier, le $[Ln(HMTDPA)_3]^{3-}$ ne cristallise pas systématiquement, il précipite aussi sous la forme d'un précipité cristallin opaque et dense. Comme le complexe $[Ln(DPA)_3]^{3-}$, étant chargé trois fois négativement, le $[Ln(HMTDPA)_3]^{3-}$est toujours accompagné de trois contre-ions. Ainsi, ajouté dans la goutte à de telles concentrations, il fait augmenter très fortement la force ionique de la solution, ce qui peut expliquer son auto-cristallisation dans les milieux à fortes concentrations en sels.

Cependant, une des propriétés avantageuses du complexe $[Ln(DPA)_3]^{3-}$ est conservée par le [Ln(HMT DPA)$_3]^{3-}$ et améliorée : la possibilité d'être détecté dans les cristaux par luminescence. Six à sept fois plus luminescent que le $[Eu/Tb(DPA)_3]^{3-}$ en solution et à une même concentration, les complexes de $[Eu/Tb(HMTDPA)_3]^{3-}$ sont encore plus facilement détectables que les $[Eu/Tb(DPA)_3]^{3-}$. En effet, la fixation des complexes $[Eu(HMTDPA)_3]^{3-}$ et $[Tb(HMTDPA)_3]^{3-}$ dans les cristaux peut être facilement mise en évidence à l'aide d'une simple lampe à ultraviolets de paillasse, même si les complexes ont été

utilisés à une faible concentration initiale. Les complexes fluorescent alors respectivement dans le visible du rouge et du vert.

Notes : autant les lamelles en verre siliconées utilisées avec les plaques Hampton® laissent passer les UV, autant ce n'est pas le cas des bouchons des plaques Nextal de la marque Qiagen®. Il faut donc les dévisser et les retourner avant d'exposer la goutte aux UV. Muni d'un simple appareil photo et d'une binoculaire dans le noir, il est aussi difficile de rendre compte, à travers une photo pour un manuscrit de thèse, du contraste entre la faible luminescence émise par la solution et celle des cristaux, pourtant plus intense dans le visible.

4.3 Acquisition et traitement des données de diffraction

Comme cela a été stipulé dans l'introduction, l'ensemble des données de diffraction a été enregistré avec le rayonnement synchrotron sur les lignes de lumière MAD BM30A, ID23-2 et ID29 d*e l'ESRF*, à Grenoble et PROXIMA1, ligne de lumière du centre de rayonnement synchrotron SOLEIL. Le protocole utilisé et les critères retenus pour enregistrer des données de diffraction en vue d'un phasage par les méthode SAD et MAD sont décrits dans l'introduction de ce manuscrit.

4.3.1 Statistiques d'intégration des données pour les protéines présentant une fixation du complexe

Seuls seront présentés les résultats obtenus avec les protéines pour lesquelles j'ai constaté, *in fine*, une fixation du complexe $[Ln(HMTDPA)_3]^{3-}$, soit :

- ▷ le lysozyme de blanc d'œuf de poule
- ▷ la thaumatine de *Thaumatococcus daniellii*

Nous n'avons jamais constaté de fixation du complexe sur les protéines AvEGFP, CaMalDH, HvMalDH et PhTET1-12s. Aucune fixation du complexe n'a pu être détectée sur la protéine AfUox, protéine dont la structure avait pu être pourtant résolue avec le tris-dipicolinate d'europium utilisé en trempage par Guillaume Pompidor, au cours de ses travaux de recherche. Lorsque nous affirmons ne pas avoir constaté de fixation, c'est au regard de ces trois critères :

- ▷ l'absence de signal anomal évalué à l'étape CORRECT du programme *XDS*,
- ▷ l'absence de pic dans la carte de Patterson anomale,
- ▷ l'absence de pic suite au calcul d'une synthèse de Fourier anomale avec les phases calculées à partir du modèle correspondant.

Dans les tableaux de données qui suivent, les valeurs entre parenthèses correspondent à la tranche de plus haute résolution.

4.3.1.1 Intégration des données enregistrées sur des cristaux dérivés de lysozyme de blanc d'œuf de poule

Plusieurs jeux de données de diffraction ont peu être enregistrés à partir de cristaux dérivés de HEWL avec les complexes tris-hydroxymethyltriazoledipicolinates d'europium, de terbium et d'ytterbium. Ceux ayant abouti à la construction d'un modèle sont ceux dont les statistiques sont données dans les **tableaux 4.1, 4.2 et 4.3** suivants. Les définitions de l'ensemble des indicateurs statistiques utilisés sont données dans l'introduction du manuscrit.

	HEWL (1) Tb		**HEWL (2) Eu**
[Complexe**]**$_{t_0}$ **(mM)**	2,7		2,7
[Protéine**]**$_{t_0}$ **(mM)**	0,37		0,25
Rapport $\frac{Qté\ de\ complexe}{Qté\ de\ protéine}\ t_0$	7,3		10,8
Méthode d'obtention	co-cristallisation		co-cristallisation
Source synchrotron	ESRF		SOLEIL
Ligne de lumière	FIP-BM30A		PROXIMA1
MAD	Oui		Non
λ (Å)	1,649 ($Tb_{L_{III}}$)	0,749	0,886
Groupe d'espace	P 4_32_12	P 4_32_12	P 4_32_12
Paramètres de maille (Å)	a = 76,634, c = 38,033	a = 76,915, c = 38,123	a = 77,403, c = 38,473
Contenu unité asymétrique	1 molécule		1 molécule
Résolution (Å)	38,32 - 1,90 (2,01 - 1,90)	38,46 - 1,57 (1,66 - 1,57)	19,24 - 1,35 (1,42 - 1,35)
Nombre de réflexions uniques	9355 (1318)	16363 (2292)	26192 (3738)
$\mathbf{R}_{merge}$ **(%)**	6,4 (27,0)	5,4 (27,6)	3,6 (19,3)
$\mathbf{R}_{pim}$ **(%)**	2,5 (11,0)	2,1 (11,2)	2,1 (11,9)
$\mathbf{Sig}_{ano}$ **XDS**	4,23 (1,54)	1,15 (0,78)	1,33 (0,82)
$\mathbf{R}_{ano}$ **(%)**	10,4 (15,8)	2,3 (8,6)	2,7 (9,4)
$\frac{I}{\sigma(I)}$	9,7 (2,8)	10,5 (2,8)	12,8 (3,8)
Complétude (%)	99,9 (99,3)	99,6 (97,6)	99,7 (99,2)
Redondance	13,2 (12,9)	13,8 (12,7)	7,5 (6,8)

Tableau 4.1 – Statistiques d'intégration de données issues de cristaux dérivés HEWL + $[Ln(HMTDPA)_3]^{3-}$, $1^{ère}$ partie

	HEWL (3) Yb	
[Complexe**]**$_{t_0}$ **(mM)**	6,0	
[Protéine**]**$_{t_0}$ **(mM)**	0,37	
Rapport $\frac{Qté\ de\ complexe}{Qté\ de\ protéine}\ t_0$	16,2	
Méthode d'obtention	co-cristallisation	
Source synchrotron	SOLEIL	
Ligne de lumière	PROXIMA1	
MAD	Oui	
λ (Å)	1,385 (Yb)	0,886
Groupe d'espace	P 4_32_12	P 4_32_12
Paramètres de maille (Å)	a = 77,274, c = 38,305	a = 77,357, c = 38,220
Contenu unité asymétrique	1 molécule	
Résolution (Å)	18,74 - 1,98 (2,09 - 1,98)	19,11 - 1,35 (1,42 - 1,35)
Nombre de réflexions uniques	8121 (964)	25738 (3600)
$\mathbf{R}_{merge}$ **(%)**	4,8 (10,6)	5,0 (20,7)
$\mathbf{R}_{pim}$ **(%)**	1,9 (5,7)	2,9 (12,0)
$\mathbf{Sig}_{ano}$ **XDS**	5,68 (2,66)	1,53 (0,87)
$\mathbf{R}_{ano}$ **(%)**	9,4 (12,6)	4,2 (12,0)
$\frac{I}{\sigma(I)}$	11,3 (6,4)	8,4 (3,7)
Complétude (%)	95,8 (82,1)	99,0 (96,9)
Redondance	12,7 (7,3)	6,9 (6,7)

Tableau 4.2 – Statistiques d'intégration de données issues de cristaux dérivés HEWL + $[Ln(HMTDPA)_3]^{3-}$, $2^{ème}$ partie

	HEWL (4) Eu	**HEWL (5) Eu**
Complexe$_{t_0}$ **(mole)**	$5{,}4.10^{-9}$	2.10^{-8}
Protéine$_{t_0}$ **(mole)**	$\approx 5.10^{-10}$	$\approx 5.10^{-10}$
Rapport $\frac{Qté\,de\,complexe}{Qté\,de\,protéine}\,{}_{t_0}$	$\approx$ 10,8	$\approx$ 40
Méthode d'obtention	trempage	trempage
Source synchrotron	ESRF	
Ligne de lumière	FIP-BM30A	
MAD	Non	Non
λ (Å)	1,776 (Eu$_{L_{III}}$)	1,776 (Eu$_{L_{III}}$)
Groupe d'espace	P 4_32_12	P 4_32_12
Paramètres de maille (Å)	a = 78,254, c = 37,906	a = 78,405, c = 37,702
Contenu unité asymétrique	1 molécule	1 molécule
Résolution (Å)	39,13 - 1,93 (2,04 - 1,93)	39,23 - 1,93 (2,04 - 1,93)
Nombre de réflexions uniques	9024 (1250)	9238 (1280)
R$_{merge}$ **(%)**	5,9 (14,6)	5,1 (13,9)
R$_{pim}$ **(%)**	2,4 (6,1)	2,1 (6,1)
Sig$_{ano}$ **XDS**	1,59 (0,78)	1,72 (0,92)
R$_{ano}$ **(%)**	3,5 (5,4)	3,4 (5,8)
$\frac{I}{\sigma(I)}$	9,8 (4,8)	10,7 (5,0)
Complétude (%)	97,7 (95,7)	99,6 (98,0)
Redondance	12,8 (12,5)	11,9 (11,5)

Tableau 4.3 – Statistiques d'intégration de données issues de cristaux dérivés HEWL + $[Ln(HMTDPA)_3]^{3-}$, $3^{ème}$ partie

Que ce soit par trempage ou en co-cristallisation, il a été possible d'obtenir des cristaux dérivés de lysozyme de blanc d'œuf de poule avec le complexe tris-hydroxymethyltriazoledipicolinate de lanthanide en l'utilisant à de faibles concentrations : de 2,7 mM à 10 mM. Les statistiques d'intégration des données de diffraction montrent, d'une part, que les cristaux n'ont pas été endommagés par l'introduction du complexe et, d'autre part, une forte fixation du complexe, compte tenu du signal anomal élevé que traduit la valeur du facteur R$_{ano}$, même pour un jeu de données enregistré hors seuil L$_{III}$ du lanthanide impliqué. En outre, on remarque que le signal anomal est nettement plus faible pour les cristaux dérivés obtenus par trempage (jeux n°4 et5) que pour ceux obtenus par co-cristallisation, bien que les données aient été enregistrées au seuil L$_{III}$ du lanthanide. Cependant, contrairement aux cristaux dérivés de HEWL obtenus par co-cristallisation avec le tris-dipicolinate de lanthanide dont le groupe d'espace (monoclinique, C2) s'était avéré différent de celui des cristaux natifs (Pompidor *et al.*, 2008), ceux dérivés avec le tris-hydroxymethyltriazoledipicolinate de lanthanide ont le même groupe d'espace que les cristaux natifs : P 4_32_12.

4.3.1.2 Intégration des données enregistrées sur des cristaux dérivés de thaumatine de *Thaumatococcus daniellii*

Deux jeux de données de diffraction ont pu être enregistrés à partir de cristaux dérivés de TdTHAU avec le complexe tris-hydroxymethyltriazoledipicolinates d'europium qui a permis le calcul de phases par la méthode de phasage MAD.

	TdTHAU (Eu)	
[Complexe**]**$_{t_0}$ **(mM)**	10	
[Protéine**]**$_{t_0}$ **(mM)**	0,6	
Rapport $\frac{Qté\ de\ complexe}{Qté\ de\ protéine}\ t_0$	16,67	
Méthode d'obtention	co-cristallisation	
Source synchrotron	ESRF	
Ligne de lumière	FIP-BM30A	
MAD	Oui	
λ (Å)	1,776 ($Eu_{L_{III}}$)	0,979 (Se_K)
Groupe d'espace	P 4_12_1 2	P 4_12_12
Paramètres de maille (Å)	a = 57,914, c = 149,972	a = 57,988, c = 150,184
Contenu unité asymétrique	1 molécule	
Résolution (Å)	45,84 - 1,93 (2,04 - 1,93)	45,90 - 1,30 (1,37 - 1,30)
Nombre de réflexions uniques	9024 (1250)	64239 (9169)
R$_{merge}$ **(%)**	3,3 (5,2)	4,9 (45,4)
R$_{pim}$ **(%)**	1,4 (2,3)	3,1 (29,2)
Sig$_{ano}$ **XDS**	3,0 (1,95)	0,90 (0,79)
R$_{ano}$ **(%)**	3,6 (4,6)	2,6 (21,2)
$\frac{I}{\sigma(I)}$	16,6 (13,2)	12,5 (1,7)
Complétude (%)	99,9 (99,7)	99,9 (99,4)
Redondance	11,5 (11,0)	6,6 (6,5)

Tableau 4.4 – Statistiques d'intégration des données de diffraction d'un cristal dérivé TdTHAU + $[Eu(HMTDPA)_3]^{3-}$

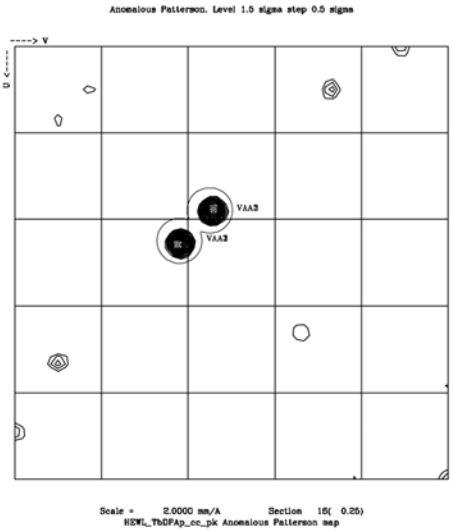
Les statistiques d'intégration de données de diffraction d'un cristal dérivé de TdTHAU montrent encore une fois une valeur assez forte du signal anomal, indiquant une fixation du complexe sur la protéine.

Contrairement au HEWL, la fixation du complexe a été mise en évidence seulement dans des cristaux dérivés par co-cristallisation. Une expérience de trempage a été menée avec une solution contenant 30 mM de complexe. Je n'ai pas détecté la fixation du complexe suite à l'analyse de l'intégration de données enregistrées au seuil d'absorption K du sélénium sur un cristal dérivé obtenu par cette méthode. Le fait de ne pas enregistrer les données au seuil L_{III} du lanthanide n'a pas permis d'avoir une valeur de f'' maximale mais dans le cas du lysozyme, cela n'avait pas empêché de mettre en évidence le signal anomal et de déterminer les phases. Le taux d'occupation du complexe introduit dans ce cristal de TdTHAU ne devait pas être élevé.

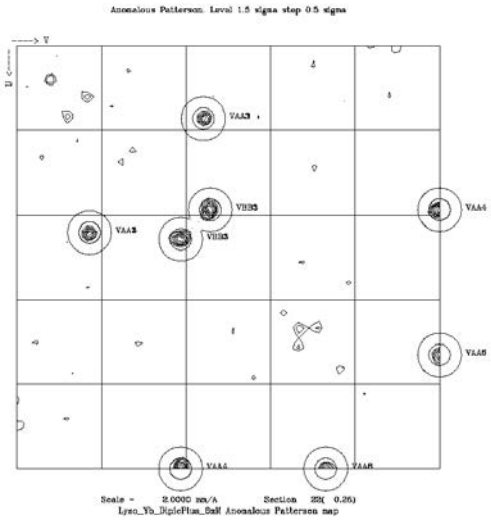
4.4 Détermination des phases expérimentales

4.4.1 Evaluation de la qualité des phases déterminées pour le lysozyme de blanc d'œuf de poule

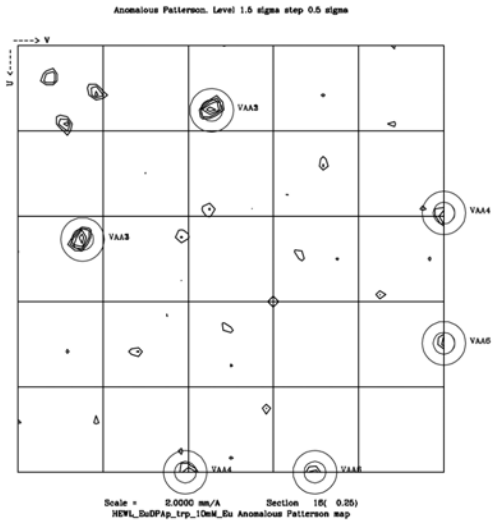
Selon le protocole décrit dans l'introduction, nous avons déterminé des jeux de phases expérimentales pour les cristaux de HEWL dérivés avec le complexe $[Eu(HMTDPA)_3]^{3-}$. La position des atomes de lanthanide a été déterminé à l'aide des cartes de Patterson anomales.


Anomalous Patterson. Level 1.5 sigma step 0.5 sigma
VAA3
VAA3
Scale = 2.0000 mm/A Section 16(0.25)
HEWL_TbDPAp_cc_pk Anomalous Patterson map

(a) Carte de Patterson anomale, section de Harker w = $\frac{1}{4}$, du cristal dérivé de HEWL avec le $[Tb(HMTDPA)_3]^{3-}$ (jeu de donnée n°1, MAD, co-cristallisation).



(b) Carte de Patterson anomale, section de Harker w = $\frac{1}{4}$, du cristal dérivé de HEWL avec le $[Yb(HMTDPA)_3]^{3-}$ (jeu de donnée n°3, MAD, co-cristallisation).



(c) Carte de Patterson anomale, section de Harker w = $\frac{1}{4}$, du cristal dérivé de HEWL avec le $[Eu(HMTDPA)_3]^{3-}$ (jeu de donnée n°5, $SAD_{Eu_{L_{III}}}$, trempage).

Figure 4.15 – Cartes de Patterson anomales calculées à partir des jeux de données de diffraction n°1, n°3 et n°5 enregistrés sur des cristaux dérivés de HEWL avec le complexe $[Ln(HMTDPA)_3]^{3-}$

La **figure 4.15** montre que le diffuseur anomal est très facilement détectable dans les cartes de Patterson anomales. La fixation du complexe se manifeste par la présence de pics intenses dans les sections de Harker (ici, la section w = $\frac{1}{4}$). Pour ces trois cartes, le plus bas niveau est de 1,5 σ et chaque pas correspond à 0,5 σ.

Au regard de l'intensité de ces pics, il est déjà possible d'avoir une idée du nombre de complexes liés à la protéine. Il est alors possible d'anticiper le succès du phasage subséquent. On remarque que dans la carte de Patterson anomale correspondant au jeu de données de diffraction issu d'un des cristaux dérivés obtenus par trempage (jeu de données n°5), l'intensité des pics est moindre par rapport aux autres cartes calculées à partir de jeux de données issus de cristaux dérivés obtenus par co-cristallisation. A partir de

cette observation, on peut déjà supposer que la la méthode de trempage utilisée ici et la concentration de $[Eu(HMTDPA)_3]^{3-}$ qu'elle impose ne permet pas d'atteindre un taux d'occupation similaire à celui constaté en co-cristallisation pour ce même site.

Les cartes de Patterson anomales calculées ont toujours mis en évidence la présence systématique de deux sites de fixation mais dont la population semblait varier, notamment pour l'un des deux sites. Nous expliquons ces différences de taux d'occupation un peu plus loin dans le chapitre (*cf.* 4.5.2.2 page 139)

Compte-tenu de la position des lanthanides dans la maille, le phasage proprement-dit a alors été entrepris. Les statistiques de phasage sont consignées dans les tableaux suivants. Comme pour celles relatives à l'intégration des données de diffraction, les définitions des indicateurs statistiques utilisés pour juger la qualité du phasage sont données dans le second chapitre de la première partie du manuscrit.

	$[Ln(HMTDPA)_3]^{3-}$		
	HEWL (1)	HEWL (2)	HEWL (3)
Rapport $\frac{Qté\,de\,complexe}{Qté\,de\,protéine}\,t_0$	7,3	10,8	16,2
Méthode d'obtention	co-cristallisation	co-cristallisation	co-cristallisation
Type de données	MAD	SAD_{rm}	MAD
Plus haute résolution (Å)	1,57	1,35	1,35
Pics$_{Fa}$ (Pic$_{Soufre}$) pk (σ)	60,85 / 23,51 (6,50)	-	45,51 / 38,13 (4,93)
Pics$_{Fa}$ (Pic$_{Soufre}$) rm (σ)	48,10 / 16,93 (4,76)	73,42 / 19,72 (4,32)	52,64 / 51,31 (4,74)
$\mathbf{FOM}^{acen.}_{sharp}$ / $\mathbf{FOM}^{cen.}_{sharp}$	0,57 / 0,46	0,40 / 0,13	0,32 / 0,16
Résolution $FOM^{acen.}_{sharp} \approx$ 0,5 (Å)	1,90	2,10	2,07
$\varphi\mathbf{p}^{\neq anomales}_{sharp\,acen.}$ / $\varphi\mathbf{p}^{\neq isomorphes}_{sharp\,acen.}$	5,31 / 3,80	1,41 / -	5,11 / 1,35
$\mathbf{SIGMAA}_{solomon}$	0,63	0,60	0,61
Arp/Warp (construit / total)	120 / 129	109 / 129	119 / 129

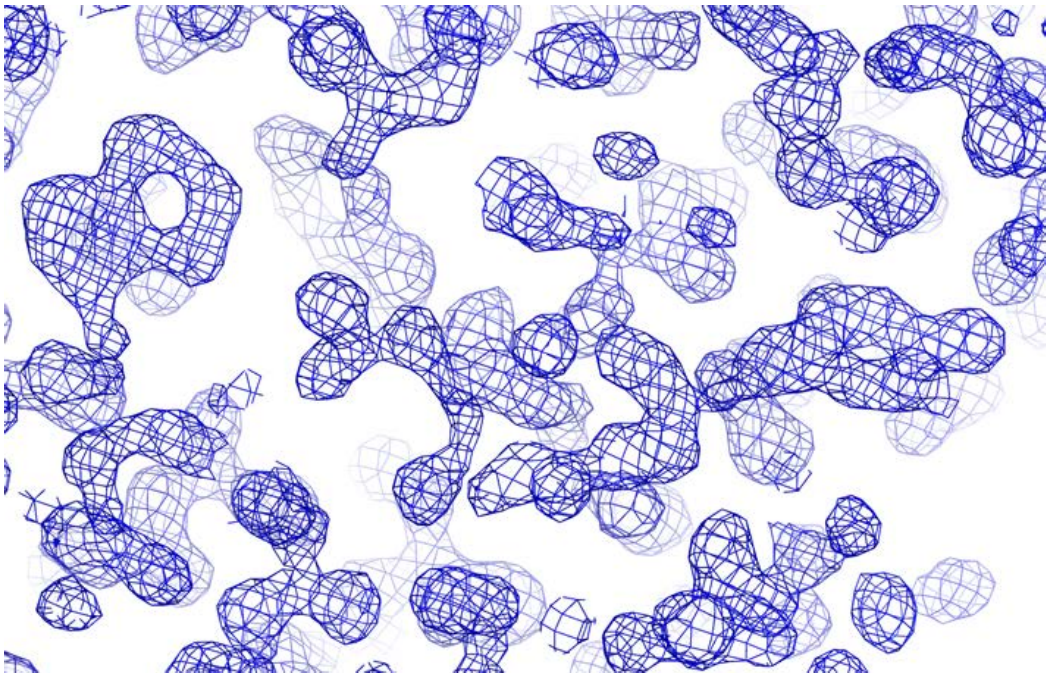
Tableau 4.5 – Statistiques de phasage expérimental pour trois jeux de données de diffraction issues de cristaux dérivés de HEWL par co-cristallisation.

	$[Ln(HMTDPA)_3]^{3-}$	
	HEWL (4)	HEWL (5)
Rapport $\frac{Qté\,de\,complexe}{Qté\,de\,protéine}\,t_0$	≈ 10,8	≈ 40
Méthode d'obtention	trempage	trempage
Type de données	$SAD_{Eu\,L_{III}}$	$SAD_{Eu\,L_{III}}$
Plus haute résolution (Å)	1,90	1,90
Pics$_{Fa}$ (Pic$_{Soufre}$) pk (σ)	30,80 / 27,45 (6,46)	29,08 / 27,16 (6,78)
Pics$_{Fa}$ (Pic$_{Soufre}$) rm (σ)	-	-
$\mathbf{FOM}^{acen.}_{sharp}$ / $\mathbf{FOM}^{cen.}_{sharp}$	0,42 / 0,02	0,43 / 0,03
Résolution $FOM^{acen.}_{sharp} \approx$ 0,5 (Å)	3,27	2,80
$\varphi\mathbf{p}^{\neq anomales}_{sharp\,acen.}$ / $\varphi\mathbf{p}^{\neq isomorphes}_{sharp\,acen.}$	1,36 / -	1,63 / -
$\mathbf{SIGMAA}_{solomon}$	0,58	0,56
Arp/Warp (construit / total)	108 / 129	58 / 129

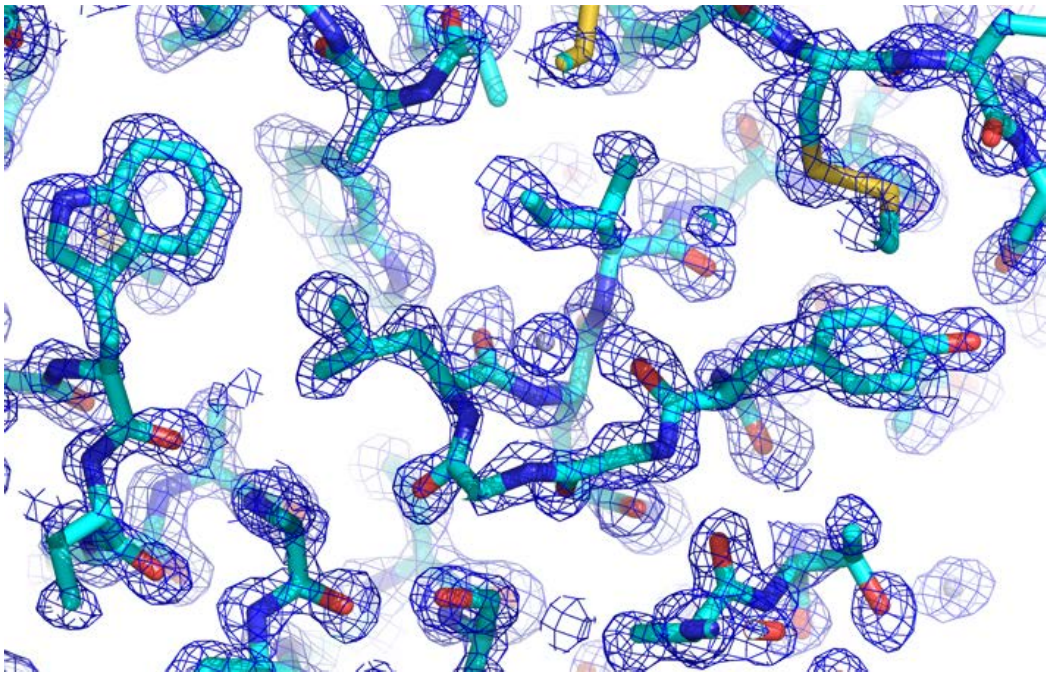
Tableau 4.6 – Statistiques de phasage expérimental pour deux jeux de données de diffraction issus de cristaux dérivés de HEWL par trempage.

Les **tableaux 4.5** et **4.6** montrent le fort pouvoir phasant du complexe $[Ln(HMTDPA)_3]^{3-}$, et ce, même lorsqu'il est utilisé à basse concentration. Qu'il s'agisse d'un phasage MAD ou même d'un phasage SAD pour des données enregistrées hors des seuils d'absorbtion L_{III} des lanthanides, les cartes de densité électronique obtenues ont conduit à la construction et à l'affinement automatique de plus de 83 % des acides aminés qui composent la protéine **[Figure 4.16]**.

Cependant, ces tableaux mettent aussi en évidence l'importance de l'information MAD sur l'efficacité du phasage **[Tableau 4.5]**. Le fait que l'écart entre la valeur de l'intensité des pics calculés par synthèse de Fourier anomale et la valeur des taux d'occupation ne varie pas de façon proportionnelle peut s'expliquer par le fait que le calcul de la synthèse de Fourier est dépendant de la résolution des données.



(a) Carte expérimentale sans modèle



(b) Carte expérimentale avec modèle affiné superposé

Figure 4.16 – Carte de densité électronique (contourée à 1,5 σ) **expérimentale** améliorée par aplatissement de solvant obtenue à partir d'un cristal dérivé de HEWL avec le $[Yb(HMTDPA)_3]^{3-}$. **-** Résolution : 1,35 Å. **Phasage SAD hors seuil** (λ = 0,886 Å)

Ces tableaux montrent aussi que le $[Ln(HMTDPA)_3]^{3-}$, introduit par trempage, conduit à des phases moins précises que lorsqu'il est introduit par co-cristallisation **[Tableau 4.6]**. Comme le montrent les statistiques d'intégration, les données enregistrées à partir de ces cristaux dérivés obtenus par trempage sont d'excellente qualité **[Tableau 4.3 page 119]**. L'obtention de cartes de densité électronique expérimentales moins interprétables ne peut donc se justifier par la qualité des données enregistrées. En trempage, les sites des diffuseurs anomaux doivent être moins peuplés. Cette observation coïncide

avec celle issue des cartes de Patterson anomales à propos des populations des sites des lanthanides qui semblent varier davantage lorsque les complexes sont introduits par trempage.

4.4.2 Evaluation de la qualité des phases déterminées pour la thaumatine de *Thaumatococcus daniellii*

Le **tableau 4.7** montre que, malgré l'enregistrement d'un jeu de données de diffraction d'excellente qualité **[Tableau 4.4 page 120]** et le calcul d'un pic de 93 σ par synthèse de Fourier anomale, il n'a pas été possible de calculer des phases expérimentales permettant de construire un modèle avec le complexe $[Eu(HMTDPA)_3]^{3-}$. En outre, on constate qu'il n'y a qu'un seul site de fixation pour le complexe à la surface de la protéine TdTHAU.

	$[Ln(HMTDPA)_3]^{3-}$
	TdTHAU
Rapport $\frac{Qté\ de\ complexe}{Qté\ de\ protéine}\ t_0$	16,67
Méthode d'obtention	co-cristallisation
Type de données	MAD
Résolution maximum (Å)	1,30
Pics$_{Fa}$ **(Pic**$_{Soufre}$**) pk (σ)**	93,37 (7,87)
Pics$_{Fa}$ **(Pic**$_{Soufre}$**) rm (σ)**	38,27 (6,99)
FOM$^{acen.}_{sharp}$ **/ FOM**$^{cen.}_{sharp}$	0,29 / 0,21
Résolution FOM$^{acen.}_{sharp} \approx$ **0,5 (Å)**	2,00
φp$^{\neq anomales}_{sharp\,acen.}$ **/ φp**$^{\neq isomorphes}_{sharp\,acen.}$	2,20 / 6,65
SIGMAA$_{solomon}$	0,40
Arp/Warp (construit / total)	0 / 207
Buccaneer (construit / total)	0 / 207

(a) Statistiques de phasage expérimental pour le jeu de données de diffraction issu du cristal dérivé de TdTHAU par co-cristallisation

	$[Ln(HMTDPA)_3]^{3-}$
	TdTHAU
LLG	1733
RFZ	24,4
TFZ	54,8

(b) Statistiques de phasage par remplacement moléculaire pour le jeu de données de diffraction issu du cristal dérivé de TdTHAU

Tableau 4.7 – Statistiques de phasage expérimental pour le jeu de données de diffraction issu du cristal dérivé de TdTHAU

Le phasage expérimental pour la thaumatine a échoué car le centre de l'unique complexe fixé à la protéine se situe sur un axe de symétrie cristallin d'ordre 2. De fait, le site principal du diffuseur anomal a une origine glissante sur l'axe z. Ceci est présenté sur la carte de Patterson anomale de la **Figure 4.17**. Les deux autres pics intenses correspondent aux extrémités des vecteurs inter-atomiques entre un atome de lanthanide et un atome de soufre impliqué dans un pont disulfure cassé sous l'effet des dommages d'irradiations.

Afin d'appréhender tout de même le mode de fixation du complexe $[Eu(HMTDPA)_3]^{3-}$, j'ai donc eu recours au remplacement moléculaire dont les indicateurs statistiques résultants sont présentés dans le **tableau 4.7**. Les définitions des indicateurs LLG, RFZ et TFZ sont aussi indiquées dans le premier chapitre

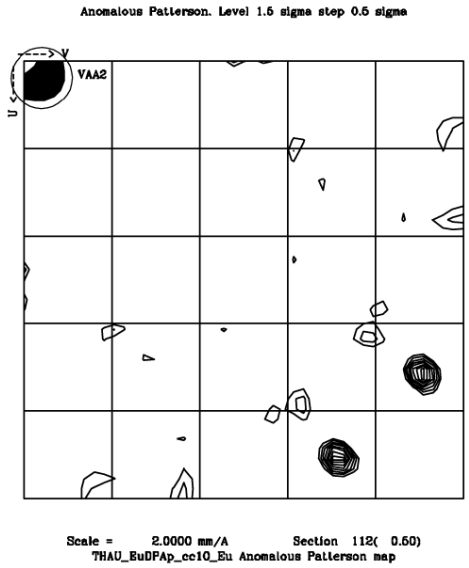

Anomalous Patterson. Level 1.5 sigma step 0.5 sigma
VAA2
U
Scale = 2.0000 mm/A Section 112(0.50)
THAU_EuDPAp_cc10_Eu Anomalous Patterson map

Figure 4.17 – Carte de Patterson anomale, section de Harker w = $\frac{1}{2}$, du cristal dérivé de TdTHAU avec le $[Eu(HMTDPA)_3]^{3-}$ (1,5 σ, pas = 0,5 σ).

4.5 Affinement des structures et description du mode de fixation du $[Ln(HMTDPA)_3]^{3-}$

Un premier modèle est obtenu par construction automatique suite au calcul d'un jeu de phases expérimentales. Nous avons ensuite complété et affiné ce modèle afin d'étudier le mode de fixation du complexe $[Ln(HMTDPA)_3]^{3-}$ sur la protéine. Contrairement aux complexes de gadolinium avec les malate déshydrogénases, je n'ai jamais été confronté à des difficultés concernant la modélisation et l'affinement du tris-hydroxymethyltriazole-dipicolinate de lanthanide lorsqu'il a permis le phasage expérimental auparavant. En effet, lorsqu'il y a eu réussite du phasage, les taux d'occupation des sites de complexe ont toujours permis d'obtenir une carte de densité électronique définie pour que je puisse modéliser les ligands des complexes. En outre, la qualité des cartes de densité électronique et les valeurs de taux d'occupation du complexe obtenues lors des premiers affinements m'ont incité à préparer des dérivés avec différentes concentrations de complexe selon les deux méthodes d'insertion : la co-cristallisation ou le trempage. Cela m'a permis d'estimer l'influence de ces différents paramètres sur l'ensemble du processus de determination de structure. Dans cette partie sont décrits les sites de fixation du complexe $[Ln(HMTDPA)_3]^{3-}$.

4.5.1 Statistiques des affinements des structures issues de cristaux dérivés

Les valeurs des différents indicateurs statistiques indiquées dans cette partie sont relatives aux affinements réalisés avec le programme Phenix et proviennent des analyses menées par le programme de pré-validation de structure *phenix.validate* de la *suite Phenix* (*cf.* introduction).

4.5.1.1 Statistiques des affinements des structures du HEWL avec le $[Ln(HMTDPA)_3]^{3-}$

Les indicateurs statistiques permettant de juger de la qualité des modèles de HEWL affinés avec les complexes $[Ln(HMTDPA)_3]^{3-}$ modélisés sont donnés dans les **tableaux 4.5 et 4.6** suivants. Comme

toujours, les définitions de l'ensemble des indicateurs statistiques utilisés ici sont consignées dans l'introduction de ce manuscrit.

	[Ln(HMTDPA)$_3$]		
	HEWL$_{co-crist.}$(1)	HEWL$_{co-crist.}$(2)	HEWL$_{co-crist.}$(3)
Rapport $\frac{Qté\,de\,complexe}{Qté\,de\,protéine}{}_{t_0}$	7,3	10,8	16,2
Code RCSB-PDB	-	4BAD	-
Résolution (Å)	38,46 - 1,57	19,74 - 1,35	19,77 - 1,35
R$_{work}$ (%)	17,99	15,27	12,40
R$_{free}$ (%)	19,86	17,19	16,32
Nbre de réflexions utilisées	30113	26140	47398
Composition (nbre d'atomes)			
Protéine	1001	1091	1165
H2O	142	159	217
Ions	12	11	5
Ligands	80	80	99
Res. hors Ramachandran	0	0	0
Ecarts types (globaux)			
Liaisons$_{Longueurs}$ (Å)	0,006	0,019	0,018
Liaisons$_{Angles}$ (°)	1,055	1,861	1,751
B$_{moyen}$ protéine (Å^2)	13,22	13,17	13,49
B$_{min}$ protéine (Å^2)	4,81	6,04	6,51
B$_{max}$ protéine (Å^2)	41,76	39,32	35,44
B$_{moyen}$ eau (Å^2)	26,08	26,50	30,40
B$_{moyen}$ ions lanthanide (Å^2)	16,42	17,80	11,21
B$_{moyen}$ ligands (Å^2)	16,02	18,10	11,40
Sites$_{Occ}$ (%)	70 / 30	65 / 24	48 / 47

Tableau 4.8 – Résumé des statistiques d'affinement pour les structures de lysozyme de blanc d'œuf de poule avec le $[Ln(HMTDPA)_3]^{3-}$. Cristaux dérivés obtenus par co-cristallisation.

	$[Ln(HMTDPA)_3]^{3-}$	
	$HEWL_{tremp.}$ (4)	$HEWL_{tremp.}$ (5)
Rapport $\frac{Qté\ de\ complexe}{Qté\ de\ protéine}_{t_0}$	≈ 10,8	≈ 40
Résolution (Å)	39,13 - 1,93	39,23 - 1,93
R_{work} (%)	18,62	18,76
R_{free} (%)	22,85	24,09
Nbre de réflexions utilisées	8903	9070
Composition (nbre d'atomes)		
Protéine	1009	1001
H2O	160	154
Ions	12	12
Ligands	51	23
Res. hors Ramachandran	0	0
Ecarts types (globaux)		
$Liaisons_{Longueurs}$ (Å)	0,011	0,016
$Liaisons_{Angles}$ (°)	0,984	1,530
B_{moyen} protéine (Å^2)	17,59	19,18
B_{min} protéine (Å^2)	7,37	8,51
B_{max} protéine (Å^2)	54,46	63,92
B_{moyen} eau (Å^2)	27,42	28,84
B_{moyen} ions lanthanide (Å^2)	14,52	19,21
B_{moyen} ligands (Å^2)	14,70	20,56
$Sites_{Occ}$ (%)	0 / 14	15 / 17

Figure 4.18 – Résumé des statistiques d'affinement pour les structures de lysozyme de blanc d'œuf de poule avec le $[Ln(HMTDPA)_3]^{3-}$. Cristaux dérivés par trempage.

Pour l'affinement des jeux de données à 1,35 Å, j'ai procédé à un affinement avec correction de l'anisotropie sur l'ensemble des atomes. Il en a résulté une chute remarquable du R_{work} et du R_{free}. Ceci explique l'écart entre les valeurs de leurs facteurs R et ceux des autres jeux de données.

4.5.1.2 Statistiques des affinements des structures de la TdTHAU avec le $[Ln(HMTDPA)_3]^{3-}$

Le **tableau 4.9** suivant permet de juger la qualité du modèle qui comprend la protéine thaumatine et le complexe $[Eu(HMTDPA)_3]^{3-}$ complet. Cet affinement fait suite au phasage par la méthode du remplacement moléculaire.

	$[Ln(HMTDPA)_3]^{3-}$
	$TdTHAU_{co-cristallisation}$
Rapport $\frac{Qté\,de\,complexe}{Qté\,de\,protéine}t_0$	16,67
Code RCSB-PDB	4BAL
Résolution (Å)	28,99 - 1,30
$\mathbf{R}_{work}$ **(%)**	13,56
$\mathbf{R}_{free}$ **(%)**	15,78
Nbre de réflexions utilisées	64138
Composition (nbre d'atomes)	
Protéine	1592
H2O	376
Ions	1
Ligands	57
Res. hors Ramachandran (%)	0
Ecarts types (globaux)	
$\mathbf{Liaisons}_{Longueurs}$ **(Å)**	0,009
$\mathbf{Liaisons}_{Angles}$ **(°)**	1,269
$\mathbf{B}_{moyen}$ **protéine (Å^2)**	11,90
$\mathbf{B}_{min}$ **protéine (Å^2)**	5,27
$\mathbf{B}_{max}$ **protéine (Å^2)**	38,27
$\mathbf{B}_{moyen}$ **eau (Å^2)**	28,40
$\mathbf{B}_{moyen}$ **ions lanthanide (Å^2)**	15,77
$\mathbf{B}_{moyen}$ **ligands (Å^2)**	21,17
$\mathbf{Site}_{Occ}$ **(%)**	48

Tableau 4.9 – Résumé des statistiques d'affinement pour la structure de la TdTHAU avec le $[Ln(HMTDPA)_3]^{3-}$

4.5.2 Description des modes de fixation du $[Ln(HMTDPA)_3]^{3-}$

4.5.2.1 Le $[Ln(HMTDPA)_3]^{3-}$ à la surface du lysozyme de blanc d'œuf de poule

Deux sites de fixation du complexe à la surface du lysozyme ont pu être déterminés. Les deux sites se situent à l'interface entre deux molécules de protéine dans le cristal. Le nouveau complexe $Ln(HMTDPA)_3]^{3-}$ conserve donc l'aspect supramoléculaire du $[Ln(DPA)_3]^{3-}$ mais sans engendrer un empilement cristallin différent de celui d'un cristal natif. Le complexe $Ln(HMTDPA)_3]^{3-}$ diffère du complexe $[Ln(DPA)_3]^{3-}$ non-seulement par le nombre de site de fixation à la surface du HEWL (le $[Ln(DPA)_3]^{3-}$ se fixe au lysozyme dans quatre sites distincts) mais aussi par son mode de fixation principal. En effet, la fixation du $[Ln(HMTDPA)_3]^{3-}$ à la protéine se fait essentiellement par des liaisons hydrophobes entre les cycles triazoles des ligands hydroxymethyltriazoledipicolinates et les cycles benzéniques des noyaux indoles des acides aminés tryptophane. Il s'agit de liaisons dites empilement-π (ou "π-π") et CH-π. Pour renforcer les interactions hydrophobes, on retrouve les liaisons hydrogènes entre les groupements carboxylates des ligands et des donneurs de protons de la protéine (en particulier les groupements amines), liaisons qui constituent le mode de fixation principal du $[Ln(DPA)_3]^{3-}$ (Pompidor *et al.*, 2008). Le mode de fixation du $Ln(HMTDPA)_3]^{3-}$ diffère donc de celui du complexe $[Ln(DPA)_3]^{3-}$.

Site 1, taux d'occupation maximum constaté : 70 %

Le premier site de fixation se situe à l'interface entre deux molécules de lysozyme dans l'empilement cristallin. Le premier complexe y établit donc des liaisons supramoléculaires avec deux molécules de lysozyme dans le cristal. Je décris ici le site de fixation du complexe dans un cristal de lysozyme obtenu par co-cristallisation. Dans la suite du manuscrit, il est probable que nous utilisions les termes "Site 1" ou "Site Trp123" pour qualifier ce site de fixation. L'unité de l'ensemble des distances indiquées dans les figures est l'angström.

Site 1, description de l'interaction du premier ligand avec la protéine Un groupement hydroxymethyltriazole est relié à la protéine de référence par une interaction hydrophobe CH-π entre son cycle triazole et le cycle benzénique du noyau indole du résidu Trp123 **[Figure 4.19]**. La distance mesurée entre les deux cycles est de 3,3 Å. De plus, un de ses groupements carboxylates établi une liaison hydrogène avec le groupement $N_\zeta H_2$ de la Lys33 de la molécule de référence. Une double conformation a pu être modélisée pour cette lysine qui décrit la position de sa chaine latérale en présence et en absence du complexe, leur taux d'occupation respectif étant en accord avec celui du complexe. Enfin, ce même ligand est lié à la molécule de référence par sa fonction hydroxyle et la fonction carboxyle de l'Asp119 via une molécule d'eau. Pour la suite du manuscrit, nous nommons "ligand 1", ce ligand qui établit la liaison CH-π avec la protéine.

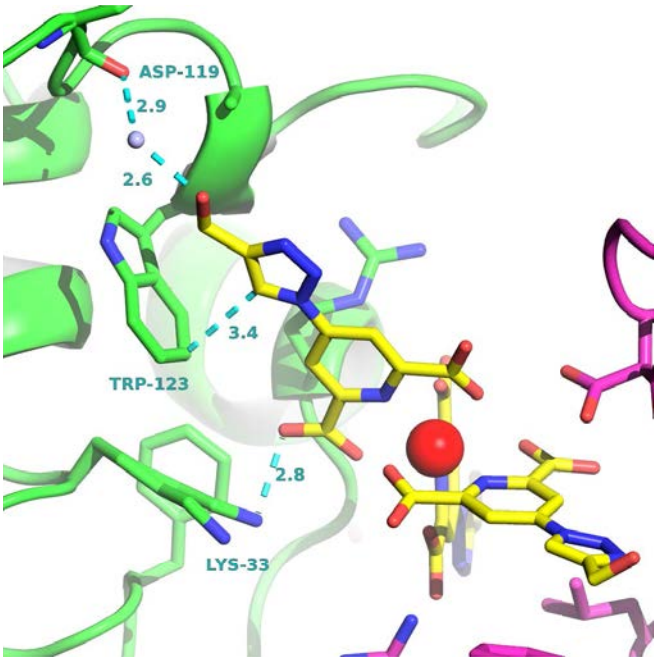


Figure 4.19 – Mode de fixation du 1^{er} ligand du complexe $[Ln(HMTDPA)_3]^{3-}$ à la surface du HEWL dans le premier site. Cristal dérivé par co-cristallisation. La molécule de référence est représentée en vert, la molécule symétrique en rose

Site 1, description de l'interaction du second ligand avec la protéine Le deuxième ligand ("ligand 2") est lié par une interaction hydrophobe, de type empilement-π entre son cycle triazole et le cycle benzénique du noyau indole du Trp62 de la molécule symétrique **[Figure 4.20]**. La distance mesurée entre les deux cycles est de 3,4 Å. Bien que la chaine latérale de l'Arg73 de la molécule

symétrique soit orientée vers le centre du complexe, sa distance de 3,9 Å vis-a-vis d'un des groupements carboxylates de ce ligand est trop grande pour permettre une liaison hydrogène. L'Arg73 interagit avec le troisième ligand. Pour la suite du manuscrit, nous nommons "ligand 2", ce ligand qui établit la liaison π-π avec la protéine.

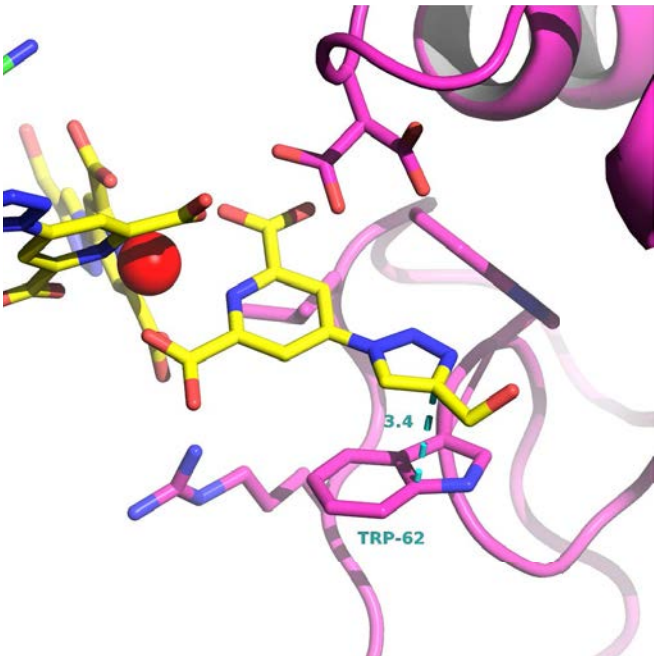


Figure 4.20 – Mode de fixation du 2^d ligand du complexe $[Ln(HMTDPA)_3]^{3-}$ à la surface du HEWL dans le premier site. Cristal dérivé par co-cristallisation. La molécule de référence est représentée en vert, la molécule symétrique en rose

Site 1, description de l'interaction du troisième ligand avec la protéine Le troisième ligand ("ligand 3") est fixé à la molécule de référence par une liaison hydrogène entre un de ses groupements carboxylates et le groupement $N_\varepsilon H$ de l'Arg5 **[Figure 4.21]**. Enfin, la distance de 3,1 Å entre une des fonctions carboxylates de ce ligand et le groupement NH1 de l'Arg73 de la molécule symétrique montre une liaison hydrogène supplémentaire entre ces deux entités.

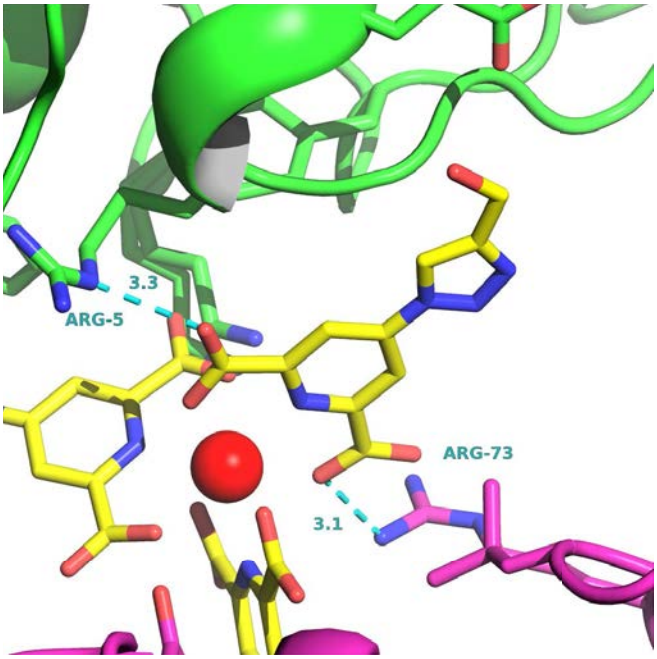


Figure 4.21 – Mode de fixation du $3^{ème}$ ligand du complexe $[Ln(HMTDPA)_3]^{3-}$ à la surface du HEWL dans le premier site. Cristal dérivé par co-cristallisation. La molécule de référence est représentée en vert, la molécule symétrique en rose

La comparaison de cette structure à celle de HEWL résolue avec le complexe $[Ln(DPA)_3]^{3-}$ révèle que ce site de fixation principal du $[Ln(HMTDPA)_3]^{3-}$ se situe à 9,7 Å du site de fixation principal du $[Ln(DPA)_3]^{3-}$ le plus proche (occupé à 88 %). Aussi, tout comme le complexe $[Ln(DPA)_3]^{3-}$ fixé à la même protéine, le $[Ln(HMTDPA)_3]^{3-}$ vient se fixer dans une zone chargée positivement. Les acides aminés donneurs de protons (arginine et lysine en l'occurrence) étant impliqués dans des liaisons hydrogènes avec les fonctions carboxylates de la partie dipicolinique des ligands **[Figure 4.22]**.

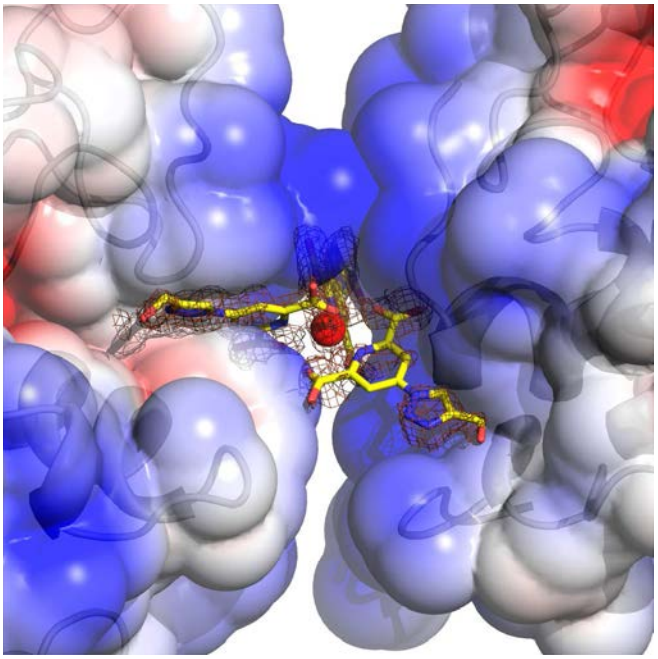


Figure 4.22 – Carte de densité électronique $2F_o$-F_c (contourée à 1,0 σ) du complexe $[Ln(HMTDPA)_3]^{3-}$et surface électrostatique autour du site de fixation principal de la protéine HEWL. La surface a été calculée à l'aide du programme APBS sur une échelle allant de -3,0 $kT.e^{-1}$ (charge négative, rouge) à +3,0 $kT.e^{-1}$ (charge positive, bleu)

Concernant ce premier site dans le cas des cristaux dérivés obtenus par trempage, que le cristal dérivé ait été préparé en présence de 2,7 mM_{t_0} ou 10 mM_{t_0} de complexe, le trop faible taux d'occupation de ce site n'a pas permis d'observer la densité électronique précise des trois ligands. Seul le ligand 1, responsable de l'interaction CH-π a pu être placé mais sa densité électronique n'a pas permis de conclure quand à la présence de l'interaction CH-π.

Site 2, taux d'occupation maximum constaté : 47 %

Quelle que soit la méthode d'obtention des cristaux dérivés de HEWL avec le complexe (co-cristallisation, trempage, concentrations en complexe), ce deuxième site est toujours moins occupé que le premier. Dans ce site, le complexe est aussi fixé à deux molécules de lysozyme, piégé dans l'empilement cristallin **[Figure 4.25]**. La plupart du temps, les taux d'occupation pour ce site m'ont seulement conduit à la modélisation d'un seul ligand voire de l'ion lanthanide seul. Un seul jeu de données (celui avec le $[Yb(HMTDPA)_3]^{3-}$ m'a permis d'aboutir à la modélisation de deux des trois ligands pour ce second complexe. J'ai pu modéliser deux ligands car les conditions de préparation des cristaux dérivés m'ont permis d'aboutir à des taux d'occupation équilibrés entre les deux sites : 48 % pour le premier site et 47 % pour le second dans les figures utilisées ici. En fin d'affinement, pour les autres modèles, les taux d'occupation des sites ont plutôt été en moyenne de 65 % pour le premier et 25 % pour le second. Dans la suite du manuscrit, nous utiliserons les termes "Site 2" ou "Site Trp62" pour qualifier ce site de fixation. L'unité de l'ensemble des distances indiqués dans les figures est l'angström.

Site 2, description de l'interaction du premier ligand avec la protéine Le premier ligand, que je nomme ainsi car il s'agit de celui que j'ai pu modéliser le plus souvent, fixe un second complexe à la protéine à l'aide de deux interactions. La première est une interaction π-π entre les atomes du noyau

triazole du ligand et les atomes N_ε, C_ζ, N1 et N2 de l'Arg112 de la molécule de référence **[Figure 4.23]**. En effet, l'extrémité de la chaine latérale de l'Arg112 se plie pour épouser la forme plane imposée par les cycles qui composent le ligand (distance de 3,5 Å). On observe le même type d'interactions entre les atomes de la fonction carboxylate de ce ligand et les atomes N_ε, C_ζ, N1 et N2 de l'Arg125 de la molécule symétrique (distance de 3,4 Å). Enfin, une liaison hydrogène s'établit entre l'autre groupement carboxylate et la fonction $C_\delta H_2$ de l'acide aminé Asn103 de la molécule de référence.

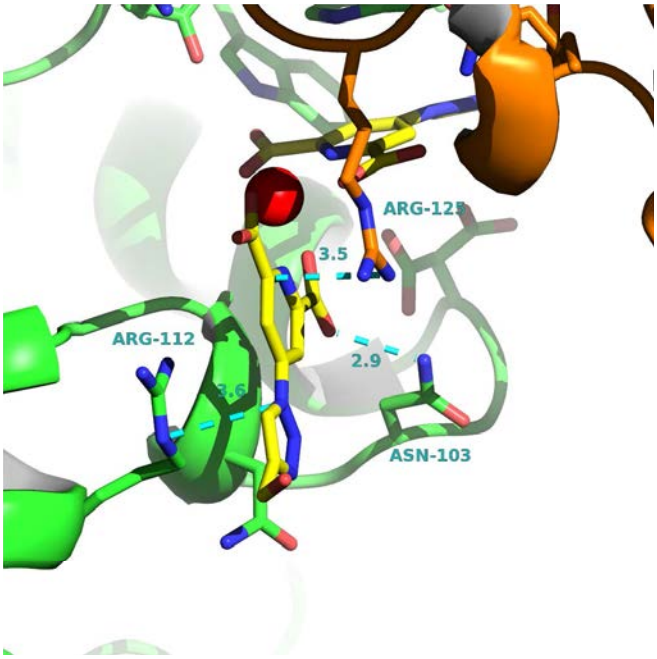


Figure 4.23 – Mode de fixation du 1^{er} ligand du complexe $[Ln(HMTDPA)_3]^{3-}$ à la surface du HEWL dans le second site. Cristal dérivé par co-cristallisation. La molécule de référence est représentée en vert, la molécule symétrique en orange

Site 2, description de l'interaction du second ligand avec la protéine Le second ligand se lie à la protéine en créant un empilement-π entre son noyau triazole et le cycle benzénique du noyau indole du Trp62 (distance de 3,4 Å) **[Figure 4.24]**. Deux liaisons hydrogènes supplémentaires viennent le stabiliser. La première se crée entre un groupement carboxylate du ligand et le groupement $N_\varepsilon H$ du résidu Trp63 de la molécule de référence. L'autre liaison hydrogène s'établit entre l'autre groupement carboxylate du ligand et l'azote $N_\varepsilon H$ de l'Arg125 de la molécule symétrique.

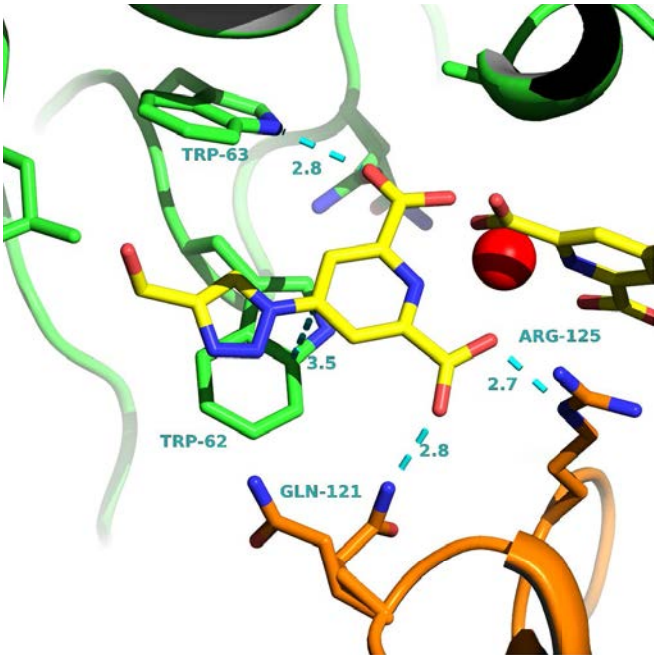


Figure 4.24 – Mode de fixation du 2^d ligand du complexe $[Ln(HMTDPA)_3]^{3-}$ à la surface du HEWL dans le second site. Cristal dérivé par co-cristallisation. La molécule de référence est représentée en vert, la molécule symétrique en orange

Comparer notre structure à celle de HEWL résolue avec le $[Ln(DPA)_3]^{3-}$ indique que ce second site de fixation du $[Ln(HMDPA)_3]^{3-}$correspond au troisième site de fixation du $[Ln(DPA)_3]^{3-}$ où un des ligands du complexe $[Ln(DPA)_3]^{3-}$établit aussi une liaison π-π entre le cycle de l'acide dipicolinique et le cycle benzénique du résidu Trp62. Dans le cas du $[Ln(DPA)_3]^{3-}$, le taux d'occupation maximal constaté pour ce site est de 62 %. Mais l'interaction hydrophobe π-π se faisant cette fois avec le cycle triazole du complexe $[Ln(HMTDPA)_3]^{3-}$, le centre du complexe $[Ln(HMTDPA)_3]^{3-}$ est décalé de 5,0 Å par rapport à celui du $[Ln(DPA)_3]^{3-}$.

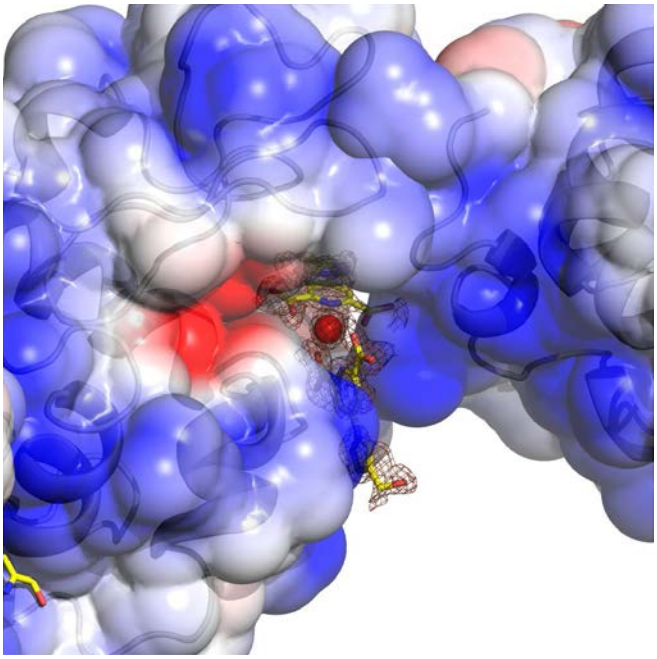


Figure 4.25 – Carte de densité électronique $2F_o$-F_c (contourée à 1,0 σ) du complexe $[Ln(HMTDPA)_3]^{3-}$ et surface électrostatique autour du second site de fixation sur la protéine HEWL. La surface a été calculée à l'aide du programme APBS sur une échelle allant de -3,0 $kT.e^{-1}$ (charge négative, rouge) à +3,0 $kT.e^{-1}$ (charge positive, bleu)

En trempage, la population du site 2 est privilégiée. J'ai ainsi pu modéliser les deux ligands décrits ci-dessus dans un cristal dérivé préparé en présence de 2,7 mM_{t_0} ou 10 mM_{t_0} de complexe. Aucune variation dans ses liaisons à la protéine n'a été constatée.

Dans le cas du lysozyme, on constate qu'aucun des deux sites de fixation du complexe $[Ln(HMTDPA)_3]^{3-}$ ne peut être peuplé à 100 % étant donné qu'ils se partagent l'interaction avec l'acide aminé Trp62.

Le complexe $[Ln(HMTDPA)_3]^{3-}$ dans l'empilement cristallin du HEWL

Je présente ici deux vues de l'empilement cristallin dans un cristal de lysozyme de groupe d'espace $P4_32_12$.

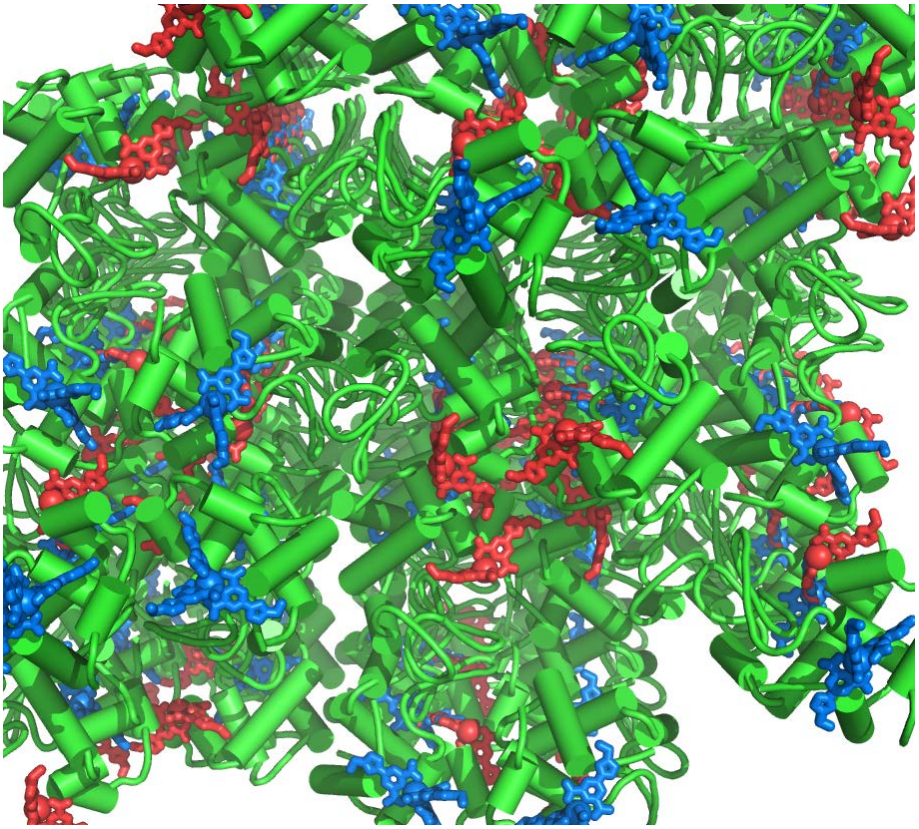


Figure 4.26 – Vue générale de l'empilement cristallin d'un cristal dérivé de lysozyme avec le complexe $[Ln(HMTDPA)_3]^{3-}$ dans une première orientation.Les molécules de protéines sont en vert, les complexes occupant les sites 1, en bleu et les complexes occupant les sites 2 sont en rouge.

Cette vue de l'empilement cristallin d'un cristal dérivé de HEWL montre la forte compacité de cet empilement (27 % de solvant). La figure met aussi en évidence que les positions des sites de fixation 1 et 2 sont alternées dans le cristal le long d'un même axe **[Figure 4.26]**.

Superposition des cinq structures de lysozyme obtenues à l'aide du $[Ln(HMTDPA)_3]^{3-}$

A l'aide de la superposition des cinq structures de lysozyme ayant pu être résolues grâce au complexe $[Ln(HMTDPA)_3]^{3-}$, il est possible d'avoir accès à une information supplémentaire relativement statistique concernant la position préférée de chaque ligand du complexe au sein d'un site de fixation. Aussi, cette méthode de représentation permet de comparer rapidement ce qui est observé suite à un trempage ou une co-cristallisation.

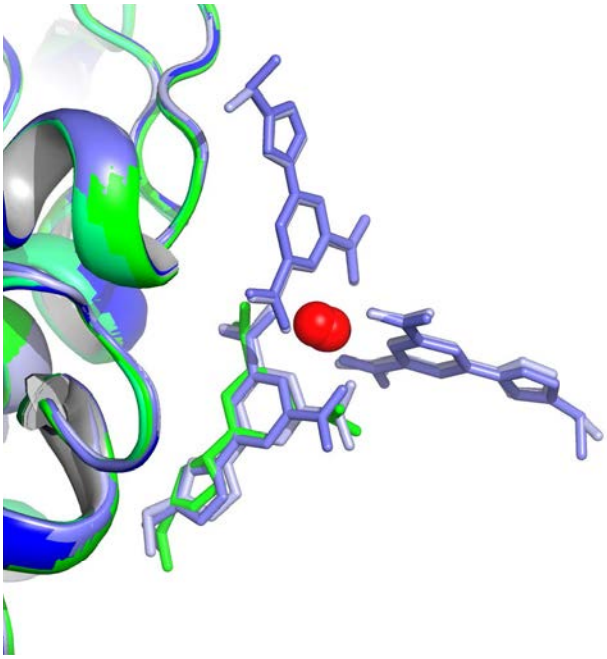


Figure 4.27 – Vue du site de fixation 1 du $[Ln(HMTDPA)_3]^{3-}$ sur la protéine HEWL. Superposition de 5 structures : 3 issues de la co-cristallisation (dominantes bleues), 2 du trempage (dominantes vertes)

On voit ainsi que quelle que soit la méthode employée, au niveau du site 1, le complexe se fixe toujours de la même façon. Les liaisons hydrophobes CH-π et π-π sont maintenues quoi qu'il advienne, seule la fonction hydroxyle adopte différentes conformations, ce qui prouve qu'elle n'intervient pas de façon indispensable dans la fixation du complexe à la protéine **[Figure 4.27]**.

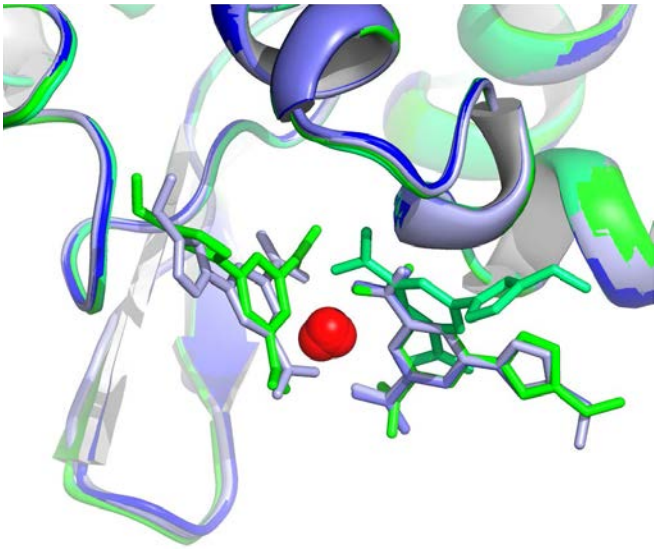


Figure 4.28 – Vue du site de fixation 2 du $[Ln(HMTDPA)_3]^{3-}$ sur la protéine HEWL. Superposition de 5 structures : 3 issues de la co-cristallisation (dominantes bleues), 2 du trempage (dominantes vertes)

Par contre, on voit que la population du second site est un peu moins assujétie à une règle systématique. La position du centre du complexe (l'ion lanthanide) est toujours respectée mais la position et la conformation des ligands qui l'entourent peuvent différer.

4.5.2.2 La corrélation entre la méthode d'obtention de cristaux dérivés, le faciès cristallin et le mode de fixation

Afin de chercher à atteindre un taux d'occupation des complexes de 100 % dans l'un des sites décrits ci-dessus, nous avons essayé obtenir d'autres cristaux dérivés en présence de concentrations bien plus élevées de complexe. Un tel taux d'occupation n'a jamais pu être atteint, que ce soit pour le site 1 ou le site 2. Cependant, les observations des gouttes de cristallisation issues de ces différentes tentatives, corrélées avec les différentes informations extraites de chacune des étapes menant à la construction des modèles affinés, nous permettent de proposer un mécanisme de fixation pour le complexe $[Ln(HMTDPA)_3]^{3-}$. L'ensemble de ces informations sont regroupées dans les tableaux **4.10 et 4.11** suivants.

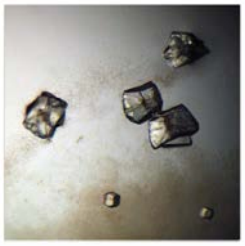
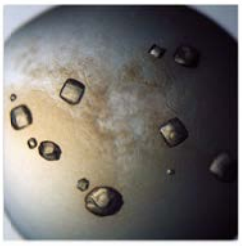
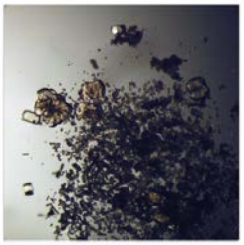
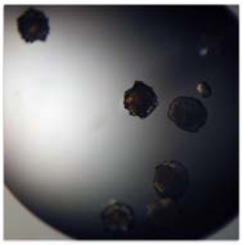
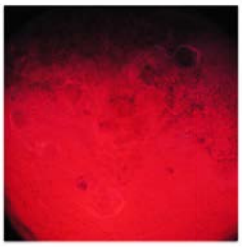
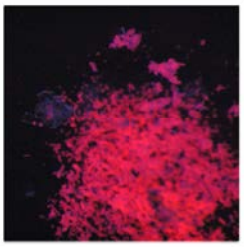
Faciès cristallin Visible				
Faciès cristallin UV				
[Complexe]$_{t0}$ [C]	2,7 mM	3,3 mM	7,1 mM	10 mM
[Protéine]$_{t0}$ [P]	0,46 mM	0,46 mM	0,70 mM	0,70 mM
[C]/[P]	5,8	7,1	10,1	14,3

Tableau 4.10 – Faciès cristallins de cristaux dérivés de HEWL par co-cristallisation avec différentes concentrations de $[Eu(HMTDPA)_3]^{3-}$

Rapport $\frac{Qté\ de\ complexe}{Qté\ de\ protéine}_{t_0}$	7,3	10,8	16,2	10,8	40
Méthode d'obtention	co-cristallisation	co-cristallisation	co-cristallisation	trempage	trempage
Site $\mathbf{1}_{Occ}$ **(%)**	70	65	48	0	15
Site $\mathbf{2}_{Occ}$ **(%)**	30	24	47	14	17
Site $\mathbf{1}_{Fa}$ **pk (σ)**	60,85	-	45,51	30,80	29,08
Site $\mathbf{2}_{Fa}$ **pk (σ)**	23,51	-	38,13	27,45	27,16
Site $\mathbf{1}_{Fa}$ **rm (σ)**	48,10	73,42	52,64	-	-
Site $\mathbf{2}_{Fa}$ **rm (σ)**	16,93	19,72	51,31	-	-

Tableau 4.11 – Méthodes d'obtention de cristaux dérivés de HEWL, taux d'occupation affinés et Fourier anomales des sites de fixation du $[Ln(HMTDPA)_3]^{3-}$

On constate que selon les méthodes de préparation de cristaux dérivés utilisées (co-cristallisation, trempage, différentes concentrations), les taux d'occupation observés varient. Ces variations s'observent déjà lors du calcul des cartes de Patterson anomales et lors du phasage, au regard de l'intensité du signal anomal mesuré pour chacun des sites de fixation.

Le **tableau 4.11** montre que la fixation du complexe est favorisée par la préparation de cristaux dérivés de HEWL par co-cristallisation plutôt que par la méthode de trempage.

Toujours dans le cas de cristaux dérivés obtenus par co-cristallisation, on voit aussi que, plus on augmente la quantité de complexe introduit dans la goutte, plus le faciès des cristaux se dégrade. Le profil de diffraction de ces cristaux au faciès torturé est de piètre qualité (par commodité, nous les avons donc qualifiés de "mauvais cristaux"). En deçà de 2,0 mM_{t_0} de complexe, la proportion de cristaux présentant un faciès propre est majoritaire. Ils ressemblent aux cristaux natifs de lysozyme **[Figure 4.7]**. Entre 2,0 mM_{t_0} et 3,3 mM_{t_0} de complexe, des cristaux aux faciès torturés apparaissent en même temps que les précédents (double faciès cristallin, voir **figure 4.8**) : c'est à partir de ce type de goutte que provient la majorité de mes jeux de données. Entre 3,3 mM_{t_0} et 10 mM_{t_0} de complexe, de plus en plus de cristaux dérivés au faciès torturé apparaissent au détriment des beaux cristaux.

De plus, qu'elle que soit la concentration en NaCl de la condition de cristallisation testée, dès que la concentration en complexe est supérieure à 10 mM_{t_0} (nous avons testé des ajouts de complexe jusqu'à plus de 30 mM), le précipité qui se forme habituellement se densifie et ne conduit plus à l'apparition de cristaux dérivés.

Comme nous l'avons vu précédemment, de par l'empilement cristallin, les sites de fixation Trp123 et Trp62 sont spatialement proches dans le cristal et forment donc une poche de fixation. Ainsi, l'addition des valeurs des taux d'occupation des deux sites par cristal est à peu près égale à 100 %. Au sein de cette poche de fixation, le complexe $[Ln(HMTDPA)_3]^{3-}$ se lie donc au Trp123 de la molécule de référence et/ou Trp62 de la molécule symétrique, conduisant à un encombrement stérique. Le taux d'occupation de l'un est dépendant de l'autre. On peut alors corréler nos observations de la manière suivante. En supposant que le complexe se lie d'abord à la protéine en solution, cela conduit à des modules protéine+complexe(s) qui vont constituer les briques de base pour la croissance du cristal. Compte tenu des observations et des explications précédentes, on peut proposer un mécanisme qui permet d'expliquer l'apparition des mauvais cristaux **[Figure 4.29]** : bien plus concentré que la protéine HEWL dans la goutte de cristallisation au départ (t_0) et de par sa plus forte affinité pour les acides aminés tryptophane, le complexe $[Ln(HMTDPA)_3]^{3-}$ va se fixer et saturer tous les sites tryptophane à

la surface de la protéine en solution. Toujours en sur-stœchiometrie par rapport à la protéine et ayant saturé tous les sites tryptophane préférentiels, les molécules de complexe libres vont chercher à se fixer à d'autres sites à la surface de HEWL. Elles vont alors utiliser les interactions de type dipicolinate pour se fixer aux arginine et lysine comme le $[Ln(DPA)_3]^{3-}$. Dès lors, le complexe $[Ln(HMTDPA)_3]^{3-}$ induit des liaisons supramoléculaires de type $[Ln(DPA)_3]^{3-}$ pour créer un nouvel empilement cristallin. Toutefois, contrairement à ce qui a été observé pour le $[Ln(DPA)_3]^{3-}$, la présence des autres complexes fixés aux acides aminés tryptophane conduit à un empilement mal organisé et donc aux cristaux présentant un faciès torturé. Ainsi plus on augmente la concentration en complexe $[Ln(HMTDPA)_3]^{3-}$, plus on obtient de mauvais cristaux. Au fur et à mesure de l'apparition de ces mauvais cristaux, la concentration en $[Ln(HMTDPA)_3]^{3-}$ dans la goutte diminue pour être finalement en sous-stœchiometrie, ce qui conduit à une fixation uniquement sur les sites tryptophane préférentiels et ce qui produit les cristaux étudiés. Ceci explique, ainsi, pourquoi j'ai toujours observé deux faciès cristallin dans la goutte. En conclusion, seuls les modules protéine+complexe(s) de composition et de forme qui conviennent pour constituer un empilement cristallin vont être "naturellement sélectionnés" dans la goutte pour former un ou plusieurs cristaux dérivés. Ainsi, seuls seront sélectionnés les modules protéine+complexe(s) compatible avec l'empilement cristallin tenant compte des deux sites de liaison (Trp123, Trp62) dans la poche de fixation. C'est pour cette raison qu'on observe une alternance des modules protéine+complexes dans la figure représentant l'empilement cristallin **[Figure 4.26 page 137]** et que les valeurs des taux d'occupation, estimées sur l'ensemble du cristal, se complètent.

Cette analyse confirme que, dans le cas du HEWL, la liaison aux acides aminés arginine et lysine est responsable de l'effet supramoléculaire des complexes dipicolinates.

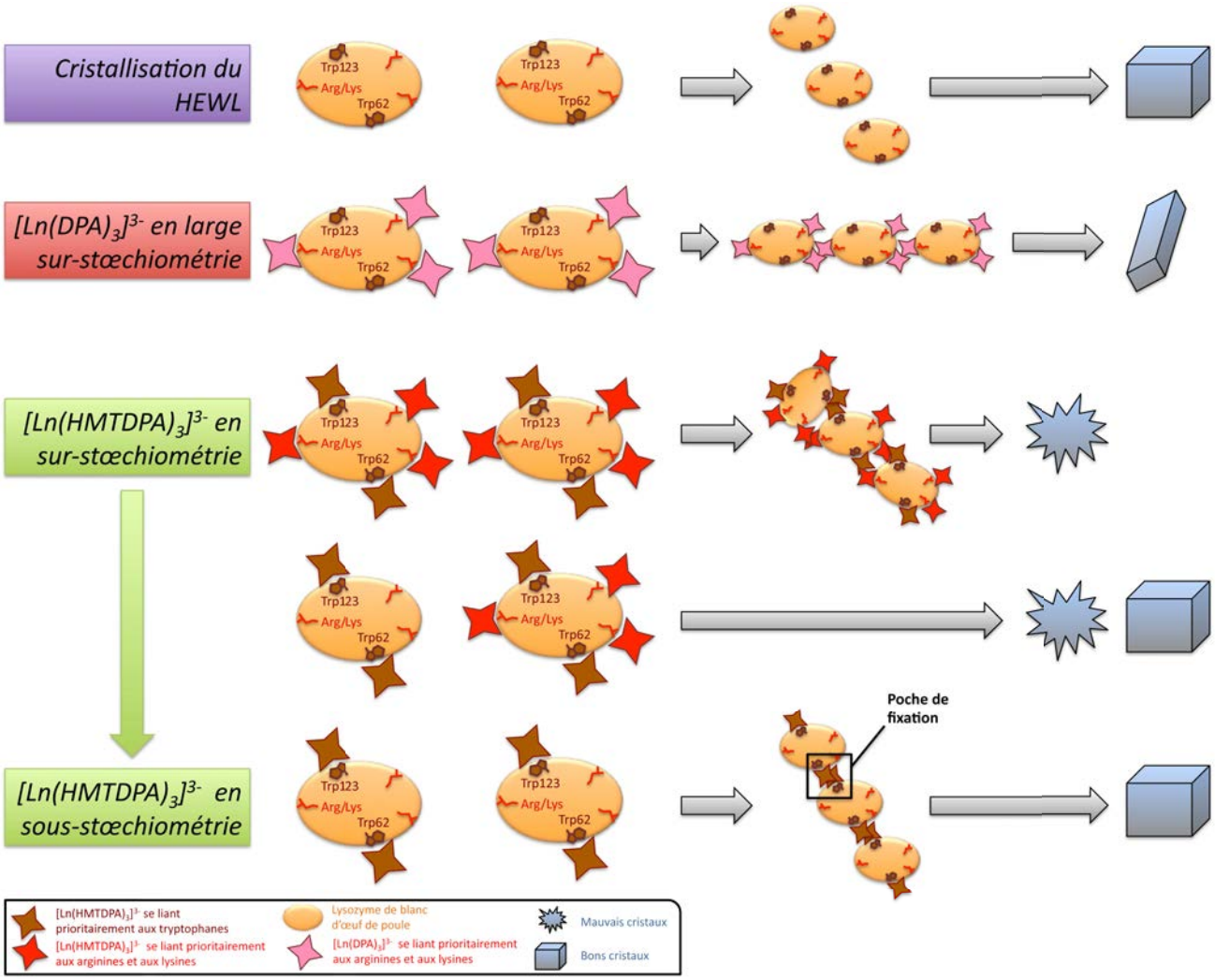


Figure 4.29 – Modèle proposé pour expliquer l'effet supramoléculaire qui conduit aux mauvais cristaux de HEWL avec le complexe $[Ln(HMTDPA)_3]^{3-}$, comparé à celui qui conduit aux cristaux monocliniques de HEWL avec le $[Ln(DPA)_3]^{3-}$

La très forte affinité du $[Ln(HMTDPA)_3]^{3-}$ pour la protéine HEWL permet de peupler aussi ces deux sites de fixation Trp62 et Trp123 par une expérience de trempage. Mais la compacité de l'empilement cristallin dans les cristaux natifs de lysozyme ne permet pas d'atteindre les taux d'occupation observés lorsque les complexes ont été introduis par co-cristallisation avec ce complexe **[Figure 4.26 page 137]**. En trempage, le second site de fixation (Trp62) a tendance à davantage se peupler que le premier site. Cela peut s'expliquer par le fait que, d'une part, les canaux de solvant des cristaux natifs de HEWL doivent proposer un accès privilégié au site Trp62 et, d'autre part, en trempage, lorsqu'un premier complexe occupe déjà la poche de fixation, il est impossible d'en fixer un deuxième. En trempage, la fixation privilégiée au site Trp62 peut donc empêcher la fixation au site Trp123, ce qui explique l'écart entre les valeurs des taux d'occupation respectifs de ces sites.

4.5.2.3 Une fixation énantiospécifique du $[Ln(HMTDPA)_3]^{3-}$

Tout comme le tris-dipicolinate de lanthanide, le tris-hydroxyméthyltriazoledipicolinate de lanthanide est un mélange racémique de deux formes énantiomères (cf. 4.1.1 page 102). L'énergie requise pour passer d'une forme à l'autre est suffisamment faible pour qu'il y ait toujours la même proportion

des deux formes en solution à température ambiante. Le phénomène d'interconversion d'une forme à l'autre est très rapide dans l'eau : de l'ordre de 10 à 100 ms (**?**). De manière générale et d'après le résultat des expériences réalisées avec le lysozyme, lorsque le taux de fixation est élevé, la densité électronique observée a toujours été sans ambiguité, elle montre clairement la fixation de l'une ou l'autre des formes énantiomères, jamais une conformation alternée des ligands. La fixation du complexe semble donc énantiospécifique. Bien évidemment, lorsque les ligands n'apparaissent pas dans les cartes de densité électronique, il est impossible de conclure quant à l'énantiospécificité de la fixation du complexe.

4.5.2.4 Le $[Ln(HMTDPA)_3]^{3-}$ à la surface de la thaumatine de *Thaumatococcus daniellii*

La structure de la protéine TdTHAU présente un seul site de fixation du complexe. Ce site est à l'interface entre deux molécules dans le cristal, centré sur un axe de symétrie d'ordre 2. Pour avoir une estimation réaliste, le taux d'occupation mesuré d'après la synthèse de Fourier (48 %) doit donc être multiplié par 2 (96 %). A nouveau, de par sa fixation à l'interface entre deux molécules, le complexe $Ln(HMTDPA)_3]^{3-}$ conserve l'aspect supramoléculaire du $[Ln(DPA)_3]^{3-}$. Le complexe se situe d'ailleurs dans la même cavité que le $[Ln(DPA)_3]^{3-}$ à la surface de la thaumatine mais il est moins enfoui. Le centre du complexe $Ln(HMTDPA)_3]^{3-}$ est décalé de 5,0 Å par rapport à celui du $[Ln(DPA)_3]^{3-}$ (Pompidor *et al.*, 2010).

Un site unique occupé à 96 % par le $[Eu(HMTDPA)_3]^{3-}$

Les interactions fines entre les ligands du complexe et la protéine sont ici plus délicates à déterminer. Le complexe se situe dans une cavité formée par l'empilement cristallin **[Figure 4.30]**.

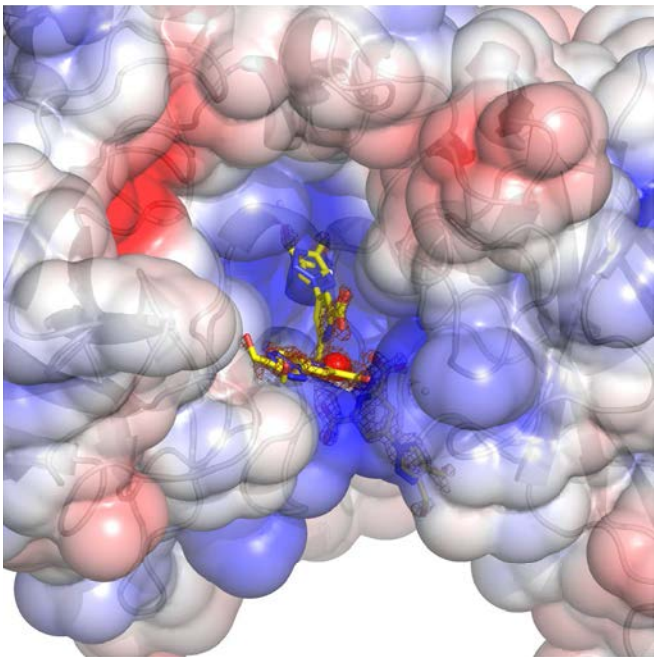


Figure 4.30 – Carte de densité électronique $2F_o$-F_c (contourée à 1,0 σ) du complexe $[Ln(HMTDPA)_3]^{3-}$ et surface électrostatique autour du site de fixation du complexe sur la protéine TdTHAU. Echelle allant de -3,0 $kT.e^{-1}$ (charge négative, rouge) à +3,0 $kT.e^{-1}$ (charge positive, bleu)

Cette cavité est chargée très positivement car on y trouve quatre acides aminés arginine (Arg79, Arg82 et leurs symétriques respectifs) et deux acides aminés lysine (la Lys78 et sa symétrique). La présence de double conformation pour la Lys78 et l'Arg79, prouve la dynamique de cette poche. Nous avons déterminé deux types d'interactions entre le complexe et la protéine **[Figure 4.31]**. La première est une liaison hydrogène entre le groupement NH_2 d'un conformère de l'Arg79 de la molécule symétrique et le groupement carboxylate du premier ligand. La seconde liaison est une interaction π-π entre les atomes du noyau triazole du ligand et les atomes N_ε, C_ζ, N1 et N2 de l'autre conformère de l'Arg79 mais cette fois de la molécule de référence (distance de 3,5 Å). Il est à noter que les taux d'occupation affinés de chacune des conformations concordent avec celui du complexe $[Eu(HMTDPA)_3]^{3-}$.

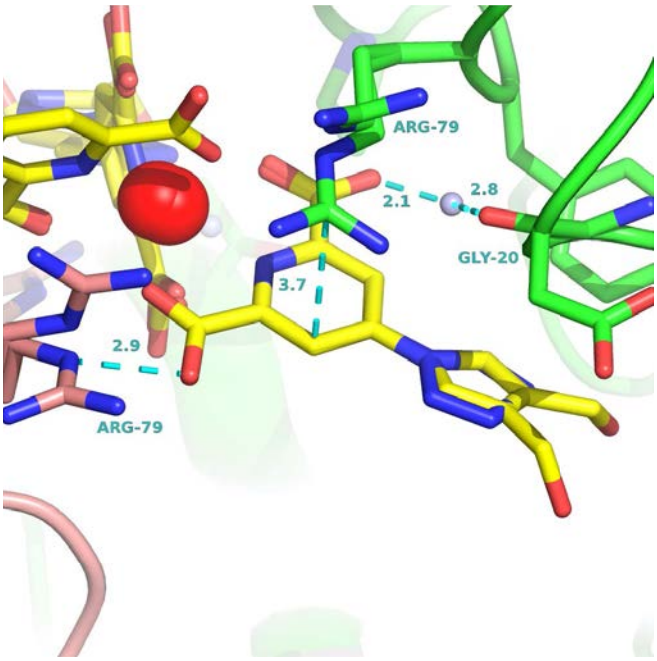


Figure 4.31 – Mode de fixation du premier ligand du complexe $[Ln(HMTDPA)_3]^{3-}$ à la surface de TdTHAU. Cristal dérivé par co-cristallisation. La molécule de référence est représentée en vert, la molécule symétrique en rose saumoné

De par l'axe de symétrie d'ordre 2, ce qui est observé pour le premier ligand l'est aussi pour le second. Ces liaisons engendrent la fermeture de la poche qui se referme alors sur le complexe en le piégeant. En outre, contrairement au lysozyme, je ne vois pas d'interaction hydrophobe de type CH-π ou π-π avec des acides aminés aromatiques. Seul l'acide aminé Phe80 orienté vers l'intérieur de la poche pourrait faire office de cible mais il est à 4,5 Å du noyau triazole du ligand. Enfin, bien que la cavité soit riche en molécules d'eau, elles ne semblent pas jouer un rôle fondamental dans la fixation des ligands. Cependant, une liaison hydrogène s'établit entre l'oxygène du ligand 1 et la fonction carboxyle de la Gly20 via une molécule d'eau (de part la symétrie, même observation pour le ligand 2).

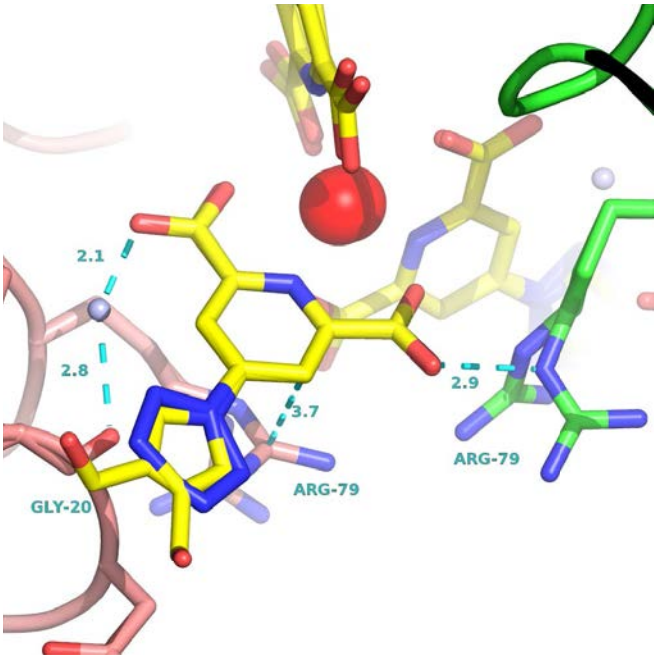


Figure 4.32 – Mode de fixation du second ligand du complexe $[Ln(HMTDPA)_3]^{3-}$ à la surface de TdTHAU. Cristal dérivé par co-cristallisation. La molécule de référence est représentée en vert, la molécule symétrique en rose saumoné

Le troisième ligand est quand à lui, placé sur l'axe de symétrie (sa densité électronique est quasiment invisible, contrairement aux deux autres ligands, voir **Figure 4.30**). Au beau milieu de la cavité, seule sa fonction hydroxyle semble établir une liaison hydrogène avec le $O_\gamma H$ de l'acide aminé Thr150 via une molécule d'eau **[Figure 4.33]**. Dans le cas de la thaumatine, les fonctions hydroxymethyles des ligands sont toujours orientées vers le solvant, au cœur de la poche.

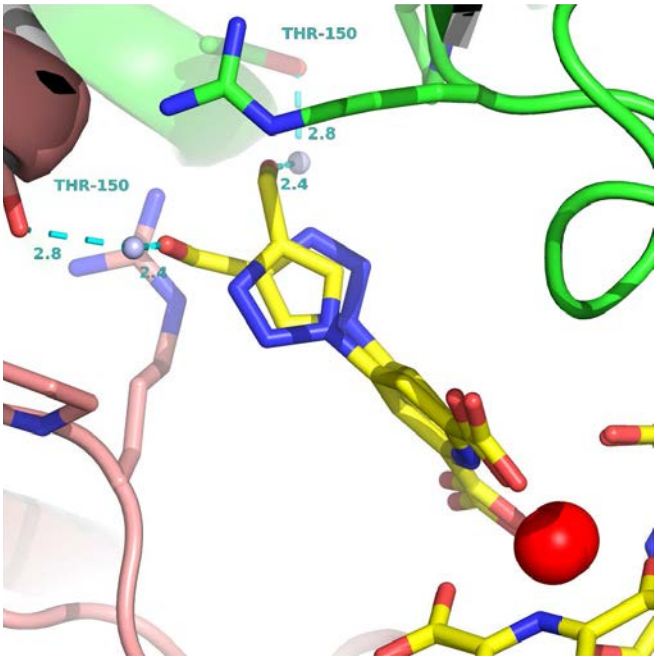


Figure 4.33 – Mode de fixation du troisième ligand du complexe $[Ln(HMTDPA)_3]^{3-}$ à la surface de TdTHAU. Cristal dérivé par co-cristallisation. La molécule de référence est représentée en vert, la molécule symétrique en rose saumoné

L'interaction du $[Ln(HMTDPA)_3]^{3-}$ avec la thaumatine se situe dans le même site que celui observé pour le $[Ln(DPA)_3]^{3-}$, à savoir une cavité formée par l'empilement cristallin et chargée positivement du au fait de la présence de nombreux acides aminés arginine. Il y a donc, dans le cas de cette protéine modèle, une interaction de type $[Ln(DPA)_3]^{3-}$. Ceci est dû au fait qu'il n'y a pas de tryptophane disponible en surface de la protéine. Le complexe $[Ln(HMTDPA)_3]^{3-}$se fixe alors via son mode de fixation secondaire : les liaisons hydrogènes avec les acides aminés donneurs de protons.

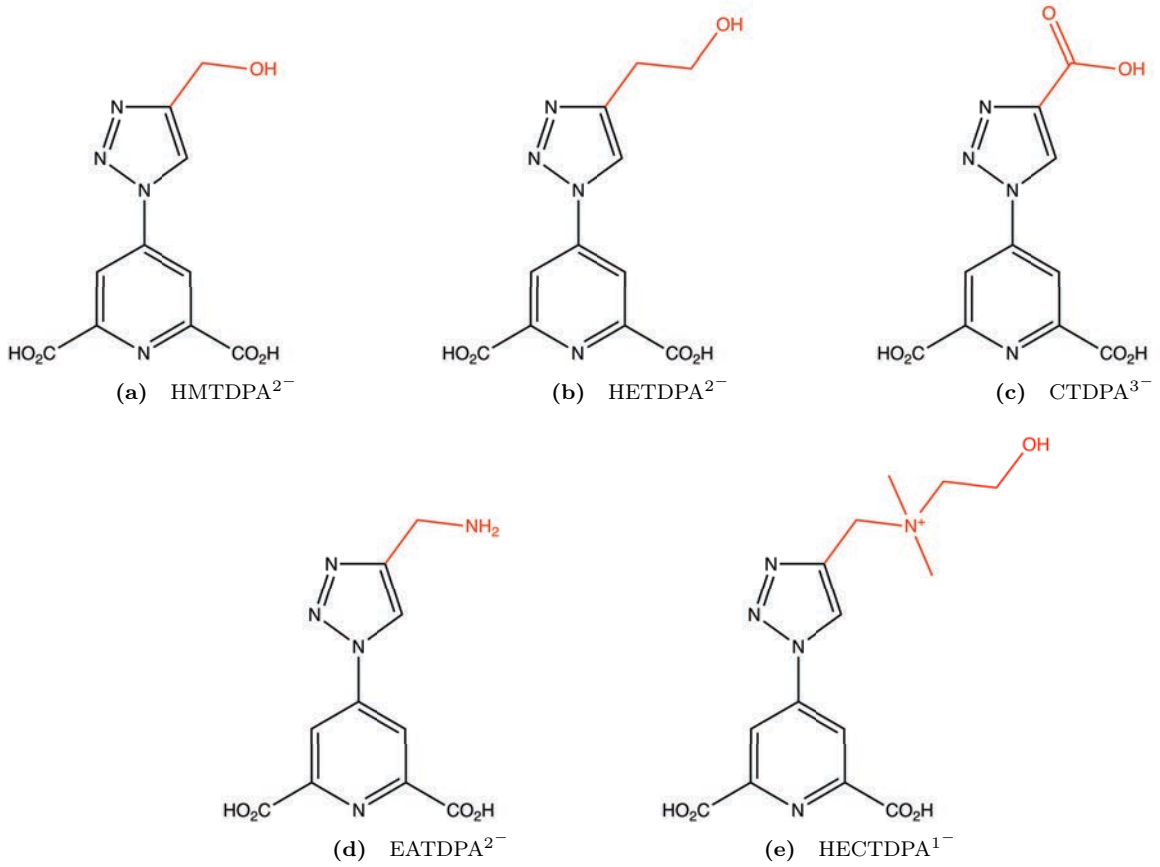
4.6 Le $[Ln(HMTDPA)_3]^{3-}$, un complexe modulable

Tout comme le tris-dipicolinate de lanthanide, le complexe tris-hydroxyméthyltriazoledipicolinate de lanthanide est formé de trois ligands identiques mais indépendants. La greffe possible par "chimie-clic" d'une fonction hydroxyméthyltriazole a donné naissance à l'idée de greffer d'autres fonctions chimiques pouvant améliorer et/ou rendre spécifique la fixation du complexe entier sur les protéines sans pour autant détériorer sa luminescence. Ainsi, quatre groupements chimiques, connus pour être impliqués dans les interactions de type liaison hydrogène du monde biologique, ont été ajoutés indépendamment au noyau triazole : une fonction amine, une fonction carboxylate, une fonction hydroxyéthyle et une fonction de type "choline" ("2-hydroxyethyl-triméthylammonium", ici 2-hydroxyethyl-diméthylammonium, la liaison à un groupement méthyle étant maintenant impliquée dans la liaison au cycle triazole. Le motif choline, omniprésent des cycles métaboliques, est en effet bien connu pour établir des liaisons fortes avec les protéines (Molina *et al.*, 2009).

Nous avons ainsi disposé de cinq ligands différents **[Figure 4.34]**. La synthèse de ces ligands revient toujours à Arnaud Gautier et Lionel Nauton, collaborateurs de l'Université Blaise-Pascal de Clermont-Ferrand, au laboratoire SEESIB.

4.6.1 Cinq ligands dipicoliniques différents pour quatre complexes différents

Les ligands $HETDPA^{2-}$, $CTDPA^{3-}$ et $EATDPA^{2-}$ nous ont été fournis séparément en poudre. Nous avons donc préparé le complexe de lanthanide selon la méthode décrite en 4.1.1. Avec ladite méthode, j'ai pu préparer en solution les complexes $[Eu(HETDPA)_3]^{3-}$ et $[Eu(CTDPA)_3]^{3-}$ à une concentration de 4 mM, que ce soit dans l'eau MilliQ ou dans une solution tampon correspondant aux conditions de cristallisation des protéines tests. Je n'ai jamais pu préparer de complexe de lanthanide avec le ligand $EATDPA^{2-}$, un précipité blanc opaque insoluble apparaissant dès l'ajout de la solution de sel de lanthanide sur celle de ligand. Il m'a été impossible de solubiliser ce précipité et ce, même après avoir soumis l'échantillon à de plus hautes températures (bain-marie 42°C et 60°C) et aux ultra-sons. Quant au ligand avec le motif choline, le $HECTDPA^{2-}$, il m'a été fourni en poudre déjà complexé avec l'europium. J'ai ainsi pu préparer une solution de $[Eu(HECTDPA)_3]$ à une concentration maximale de 35 mM.



(a) HMTDPA^{2-} **(b)** HETDPA^{2-} **(c)** CTDPA^{3-}

(d) EATDPA^{2-} **(e)** HECTDPA^{1-}

Figure 4.34 – Les différents ligands du complexe de lanthanide $[Ln(HMTDPA)_3]^{3-}$

Voici la nomenclature exacte pour chacun des ligands précités :

HMTDPA "HydroxyMethylTriazoleDipicolinate" pour "acide 4-(4-(hydroxymethyl)-1H-1,2,3-triazol-1-yl)pyridine-2,6-dicarboxylique" (308,16 g.mol^{-1})

HETDPA "HydroxyEthylTriazoleDipicolinate" pour "acide 4-(4-(2-hydroxyethyl)-1H-1,2,3-triazol-1-yl) pyridine-2,6-dicarboxylique" (322,18 g.mol^{-1}),

CTDPA "CarboxylateTriazoleDipicolinate" pour "acide 4-(4-carboxy-1H-1,2,3-triazol-1-yl)pyridine-2,6-dicarboxylique" (322,14 g.mol^{-1}),

EATDPA "EthylAmineTriazoleDipicolinate" pour "acide 4-(4-(aminomethyl)-1H-1,2,3-triazol-1-yl)pyri dine-2,6-dicarboxylique" (307,17 g.mol^{-1}),

HECTDPA "HydroxyEthylCholineTriazoleDipicolinate" pour "N-((1-(2,6-dicarboxypyridin-4-yl)-1H-1,2,3-triazol-4-yl)methyl)-2-hydroxy-N,N-dimethylethanammonium" (380,29 g.mol^{-1}).

Dans un premier temps, j'ai utilisé des solutions de complexe $[Eu(HETDPA)_3]^{3-}$, $[Eu(CTDPA)_3]^{6-}$ et $Eu(HECTDPA)_3$, respectivement concentrées à 4 mM, 4 mM et 35 mM, pour essayer d'obtenir des cristaux dérivés avec trois protéines. Les deux protéines modèles usuelles : le lysozyme de blanc d'œuf de poule et la thaumatine de *Thaumatococcus daniellii* mais j'ai aussi testé ces complexes sur la malate déshydrogénase de *Haloferax volcanii*.

Il est à noter que, tout comme le complexe $[Eu(HMTDPA)_3]^{3-}$, le complexe $[Eu(HETDPA)_3]^{3-}$ est chargé trois fois négativement, ce qui n'est plus le cas, ni du complexe $[Eu(CTDPA)_3]^{6-}$, chargé six fois négativement, ni du complexe choline $Eu(HECTDPA)_3$ dont la charge globale est nulle.

Je présente dans cette section les essais réalisés avec les complexes $[Eu(HETDPA)_3]^{3-}$, $[Eu(CTDPA)_3]^{6-}$ et $Eu(HECTDPA)_3$ ainsi que les observations associées. Nous confronterons alors les résultats obtenus avec ceux du complexe $[Eu(HMTDPA)_3]^{3-}$.

4.6.1.1 Récapitulatif des essais réalisés avec le complexe $[Eu(HETDPA)_3]^{3-}$

Obtention de cristaux dérivés de HEWL et de TdTHAU avec le $[Eu(HETDPA)_3]^{3-}$

Que ce soit au cours des différents essais de cristallisation et du phasage en lui-même, le complexe $[Eu(HETDPA)_3]^{3-}$ se comporte comme le $[Eu(HMTDPA)_3]^{3-}$. L'analyse de son mode de fixation ne montre d'ailleurs que très peu de changements par rapport à celui du $[Eu(HMTDPA)_3]^{3-}$. La taille du complexe $[Eu(HETDPA)_3]^{3-}$ correspond à une sphère de 24 Å environ (contre 22 Å pour le $[Eu(HMTDPA)_3]^{3-}$).

Comme le $[Eu(HMTDPA)_3]^{3-}$, l'introduction du complexe $[Eu(HETDPA)_3]^{3-}$ par co-cristallisation avec le lysozyme a une influence sur le faciès des cristaux dérivés obtenus. En co-cristallisation, ces cristaux dérivés sont plus petits et leurs faces sont plus torturées que celles de cristaux natifs de lysozyme. Pour les essais de co-cristallisation, j'ai pu augmenter la concentration de ce complexe dans la goutte par suspension de la protéine lyophilisée dans des solutions de complexe concentrées à 2,7 mM et 4,0 mM. Mais le complexe $[Eu(HETDPA)_3]^{3-}$ se comporte comme le $[Eu(HMTDPA)_3]^{3-}$: plus on augmente sa concentration, plus les cristaux dérivés sont torturés et plus la qualité de leur profil de diffraction baisse. Mais, à l'instar du $[Eu(HMTDPA)_3]^{3-}$, dès 2,7 mM de $[Eu(HETDPA)_3]^{3-}$, il n'est déjà plus possible d'obtenir des cristaux permettant d'enregistrer des données interprétables **[Tableau** 4.12**]**.

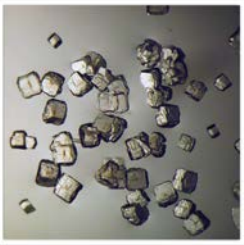
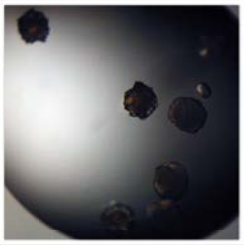
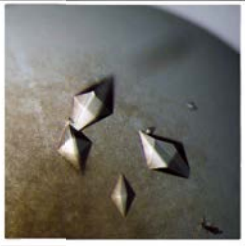
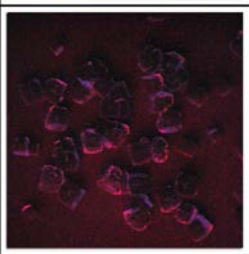
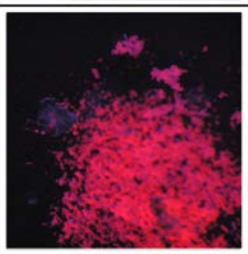
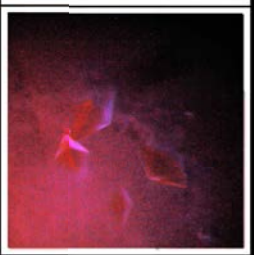
	$[Eu(HETDPA)_3]^{3-}$			
Protéine modèle	HEWL			TdTHAU
Méthode d'obtention	co-cristallisation	co-cristallisation	co-cristallisation	co-cristallisation
$[Complexe]_{t0}$ [C] (mM)	1,4	2,7	4,0	1,4
$[Protéine]_{t0}$ [P] (mM)	0,46	0,46	0,46	0,6
[C] / [P]	3,0	5,9	8,7	2,3
Faciès cristallin Visible				
Faciès cristallin UV				

Tableau 4.12 – Faciès cristallin de cristaux dérivés avec le $[Eu(HETDPA)_3]^{3-}$

On s'attend donc déjà à ce que le mode de fixation du $[Eu(HETDPA)_3]^{3-}$ soit le même que celui du $[Eu(HMTDPA)_3]^{3-}$ avec les sites tryptophane qui, une fois saturés à plus hautes concentrations, orientent le complexe vers l'utilisation des sites arginine, ce qui conduit aux mauvais cristaux.

Quand aux cristaux dérivé de thaumatine avec le $[Eu(HETDPA)_3]^{3-}$, ceux-ci conservent un faciès identique aux cristaux natifs. Là encore, le $[Eu(HETDPA)_3]^{3-}$ se comporte comme le $[Eu(HMTDPA)_3]^{3-}$ **[Tableau** 4.12**]**. Il a été aussi possible d'obtenir un cristal dérivé de HEWL avec le complexe $[Eu(HETDPA)_3]^{3-}$ par la méthode de trempage.

J'ai aussi essayé d'obtenir des cristaux dérivés de HvMalDH et de PhTET1 avec ce complexe, mais je n'ai pas constaté de fixation.

Intégration des données de diffraction issues de cristaux dérivés de HEWL avec le $[Eu(HETDPA)_3]^{3-}$

Des données MAD et SAD ont été enregistrées au seuil d'absorption L_{III} de l'europium sur des cristaux dérivés de HEWL. Les statistiques d'intégration de ces données sont répertoriées dans le tableau suivant.

	$[Eu(HETDPA)_3]^{3-}$		
	HEWL		
Rapport $\frac{Qté\ de\ complexe}{Qté\ de\ protéine}\ t_0$	3,0		≈ 10,8
Méthode	co-cristallisation		trempage
Source synchrotron	ESRF		
Ligne de lumière	FIP-BM30A		
MAD	oui		non
λ (Å)	1,776 ($Eu_{L_{III}}$)	0,979	1,776 ($Eu_{L_{III}}$)
Groupe d'espace	$P4_32_12$		$P4_32_12$
Paramètres de maille (Å)	a = 77,107, c = 37,828	a = 77,763, c = 37,896	a = 78,209, c = 38,012
Contenu unité asymétrique	1 molécule		1 molécule
Résolution (Å)	38,55 - 2,00 (2,11 - 2,00)	38,88 - 1,51 (1,59 - 1,51)	39,10 - 1,93 (2,04 -1,93)
Nombre de réflexions uniques	8077 (1118)	18899 (2644)	9278 (1304)
R_{merge} (%)	3,8 (7,1)	5,0 (32,3)	4,5 (10,6)
R_{pim} (%)	2,4 (4,7)	2,0 (13,3)	1,8 (4,5)
Sig_{ano} XDS	5,47 (3,59)	1,28 (0,81)	2,15 (1,06)
R_{ano} (%)	11,0 (15,0)	2,0 (11,0)	4,0 (5,0)
$\frac{I}{\sigma(I)}$	13,9 (9,2)	10,3 (2,4)	11,3 (6,8)
Complétude (%)	99,4 (97,1)	99,7 (98,0)	99,6 (98,4)
Redondance	6,4 (6,0)	13,6 (12,7)	12,5 (11,9)

Tableau 4.13 – Statistiques d'intégration des données de diffraction de cristaux dérivés de HEWL + $[Eu(HETDPA)_3]^{3-}$

Intégration des données de diffraction issues de cristaux dérivés de TdTHAU avec le $[Eu(HETDPA)_3]^{3-}$

Des données MAD ont été enregistrées au seuil d'absorption L_{III} de l'europium et au seuil K du sélénium sur des cristaux dérivés de TdTHAU avec le $[Eu(HETDPA)_3]^{3-}$. Les statistiques d'intégration de ces données sont indiquées dans le tableau suivant.

	$[Eu(HETDPA)_3]^{3-}$	
	TdTHAU	
Rapport $\frac{Qté\ de\ complexe}{Qté\ de\ protéine}\ t_0$	2,3	
Méthode	co-cristallisation	
Source synchrotron	ESRF	
Ligne de lumière	FIP-BM30A	
MAD	oui	
λ (Å)	1,776 ($Eu_{L_{III}}$)	0,979
Groupe d'espace	$P4_12_12$	
Paramètres de maille (Å,°)	a = 57,809, c = 150,076	a = 57,852, c = 150,215
Contenu unité asymétrique	1 molécule	
Résolution (Å)	45,79 - 1,93 (2,04 - 1,93)	45,83 - 1,20 (1,26 - 1,20)
Nombre de réflexions uniques	19052 (2614)	78373 (11008)
R_{merge} (%)	4,3 (5,7)	5,1 (45,7)
R_{pim} (%)	1,8 (2,2)	1,9 (15,7)
Sig_{ano} XDS	3,96 (3,05)	1,03 (0,79)
R_{ano} (%)	4,9 (5,8)	2,5 (15,6)
$\frac{I}{\sigma(I)}$	11,1 (11,1)	10,4 (1,7)
Complétude (%)	96,3 (93,4)	98,0 (95,8)
Redondance	13,7 (13,4)	9,9 (10,0)

Tableau 4.14 – Statistiques d'intégration des données de diffraction de cristaux dérivés de TdTHAU + $[Eu(HETDPA)_3]^{3-}$

Les statistiques issues de l'intégration des données mettent en évidence la qualité de ces dernières **[Tableaux 4.13 et 4.14]**. Elles montrent aussi qu'avec ce type de complexe, le signal anomal mesuré reste plus élevé pour un cristal dérivé par co-cristallisation que celui provenant d'un cristal dérivé par trempage **[Tableau 4.13]**.

Phasage expérimental de HEWL avec le $[Eu(HETDPA)_3]$

Avec le lysozyme de blanc d'œuf de poule, on constate que le complexe s'est suffisamment fixé dans le cristal pour permettre le phasage expérimental. Comme c'était le cas pour le complexe $[Eu(HMTDPA)_3]^{3-}$, le phasage réalisé à partir de données de diffraction issues d'un cristal dérivé par co-cristallisation est plus précis que celui réalisé à partir d'un cristal dérivé par trempage. Il permet la construction automatique de plus de 98 % des résidus qui composent la protéine **[Tableau 4.15 et Figure 4.35]**. Suite à la co-cristallisation, un seul site de fixation sert au calcul des phases alors que suite à un trempage, ces phases se calculent à partir de deux sites de fixation. En trempage, le second site de fixation se peuple davantage, au détriment du premier. D'après les coordonnées des sites de fixation des diffuseurs anomaux, on peut en déduire un comportement équivalent au $[Eu(HMTDPA)_3]^{3-}$.

	$[Eu(HETDPA)_3]^{3-}$	
	HEWL	
Rapport $\frac{Qté\ de\ complexe}{Qté\ de\ protéine}\ t_0$	3,0	$\approx$ 10,8
Méthode d'obtention	co-cristallisation	trempage
Type de données	MAD	$SAD_{Eu_{L_{III}}}$
Résolution (Å)	40,83 - 1,51	41,06 - 1,93
Pics$_{Fa}$ **(Pic**$_{Soufre}$**) pk (σ)**	58,75 (4,38)	38,38 / 22,64 (7,05)
Pics$_{Fa}$ **(Pic**$_{Soufre}$**) rm (σ)**	60,05 (5,74)	-
$\mathbf{FOM}^{acen.}_{sharp}$ / $\mathbf{FOM}^{cen.}_{sharp}$	0,44 / 0,11	0,49 / 0,33
Nombre de sites	1	2
$\mathbf{Résolution\ FOM}^{acen.}_{sharp} \approx$ **0,5 (Å)**	2,06	2,27
$\mathbf{P\varphi}^{\neq anomales}_{sharp\,acen.}$ / $\mathbf{P\varphi}^{\neq isomorphes}_{sharp\,acen.}$	6,54 / 0,80	2,51 / 0,00
$\mathbf{SIGMAA}_{solomon}$	0,58	0,55
Arp/Warp (construit / total)	127 / 129	50 / 129

(a) Statistiques de phasage expérimental

	$[Eu(HETDPA)_3]^{3-}$
	HEWL
LLG	1259
RFZ	19,3
TFZ	34,8

(b) Statistiques de phasage par remplacement moléculaire pour le jeu de données SAD

Tableau 4.15 – Statistiques de phasage issues de données de diffraction de cristaux de HEWL dérivés avec le $[Eu(HETDPA)_3]^{3-}$

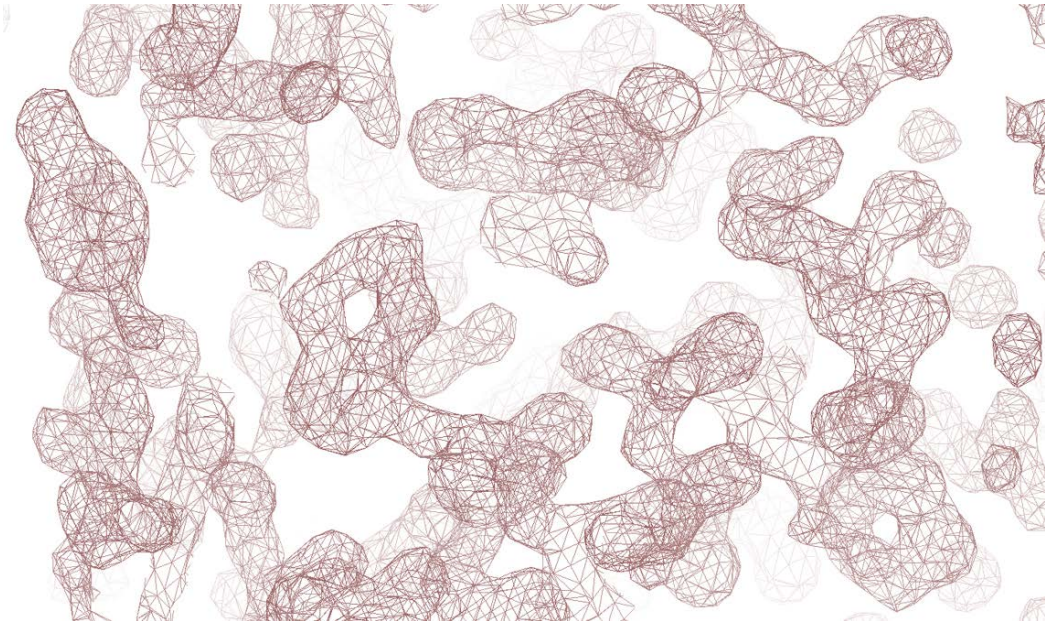


Figure 4.35 – Carte de densité électronique **expérimentale** (contourée à 1,5 σ) améliorée par aplatissement de solvant obtenue à partir d'un cristal dérivé de HEWL + $[Eu(HETDPA)_3]^{3-}$, phasage MAD, résolution : 1,20 Å

Phasage expérimental de TdTHAU avec le $[Eu(HETDPA)_3]$

Avec la thaumatine de *Thaumatococcus daniellii*, le phasage est effectué à partir du même site de fixation que celui du $[Eu(HMTDPA)_3]^{3-}$. En effet, bien que très peu éloignées, les coordonnées du lanthanide du complexe $[Eu(HETDPA)_3]^{3-}$ diffèrent très légèrement de celles qui avaient été déterminées avec le complexe $[Eu(HMTDPA)_3]^{3-}$. C'est d'ailleurs ce léger décalage qui a permis la réussite immédiate du phasage expérimental. En effet, le diffuseur anomal ne se trouvant plus strictement sur l'axe de symétrie d'ordre 2, son origine n'était plus glissante.

	$[Eu(HETDPA)_3]^{3-}$
	TdTHAU
Rapport $\frac{Qté\,de\,complexe}{Qté\,de\,protéine}{}_{t_0}$	2,3
Type de données	MAD
Résolution (Å)	48,124 - 1,20
Pics$_{Fa}$ **(Pic**$_{Soufre}$**) pk (σ)**	99,05 (6,68)
Pics$_{Fa}$ **(Pic**$_{Soufre}$**) rm (σ)**	67,73 (7,87)
FOM$^{acen.}_{sharp}$ **/ FOM**$^{cen.}_{sharp}$	0,29 / 0,17
Nombre de sites	1
Résolution FOM$^{acen.}_{sharp} \approx$ **0,5 (Å)**	2,00
P$\varphi^{\neq anomales}_{sharp\,acen.}$ **/ P**$\varphi^{\neq isomorphes}_{sharp\,acen.}$	5,02 / 1,87
SIGMAA$_{solomon}$	0,78
Arp/Warp (construit / total)	197 / 207

Tableau 4.16 – Statistiques de phasage expérimental à partir des données de diffraction de cristaux de TdTHAU dérivés avec le $[Eu(HETDPA)_3]^{3-}$

Le processus de phasage conduit à la construction automatique de plus de 95 % des acides aminés **[Tableau 4.16]**. Ce qui n'est guère étonnant au vu de la précision de la carte de densité électronique obtenue après nivelation de solvant **[Figures 4.36 et 4.37]**.

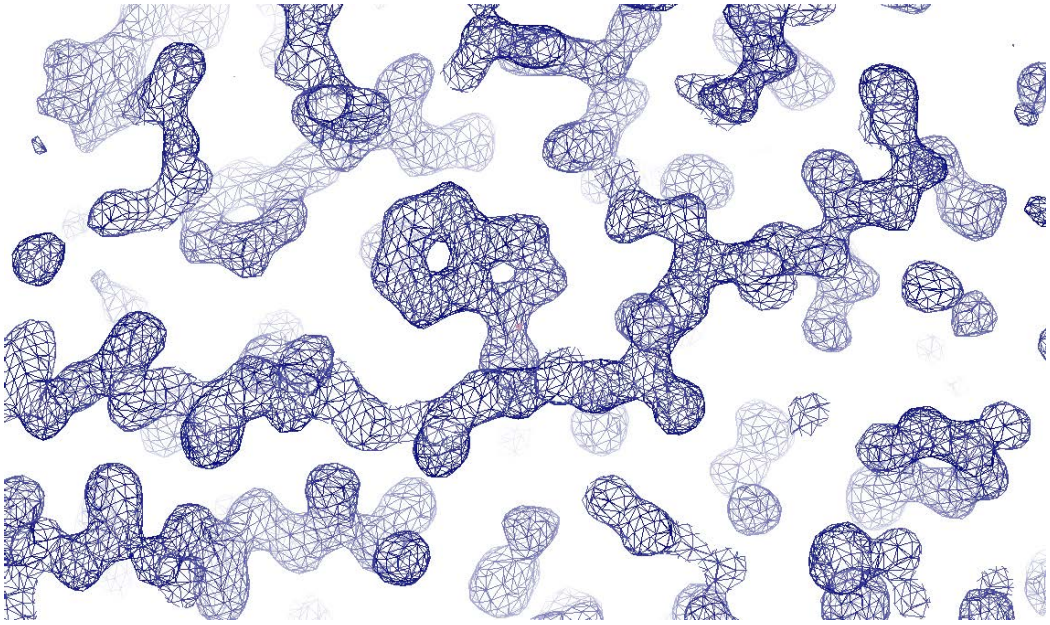


Figure 4.36 – Carte de densité électronique **expérimentale** (contourée à 1,5 σ) améliorée par aplatissement de solvant d'un cristal dérivé de TdTHAU + $[Eu(HETDPA)_3]^{3-}$, phasage MAD, résolution : 1,20 Å

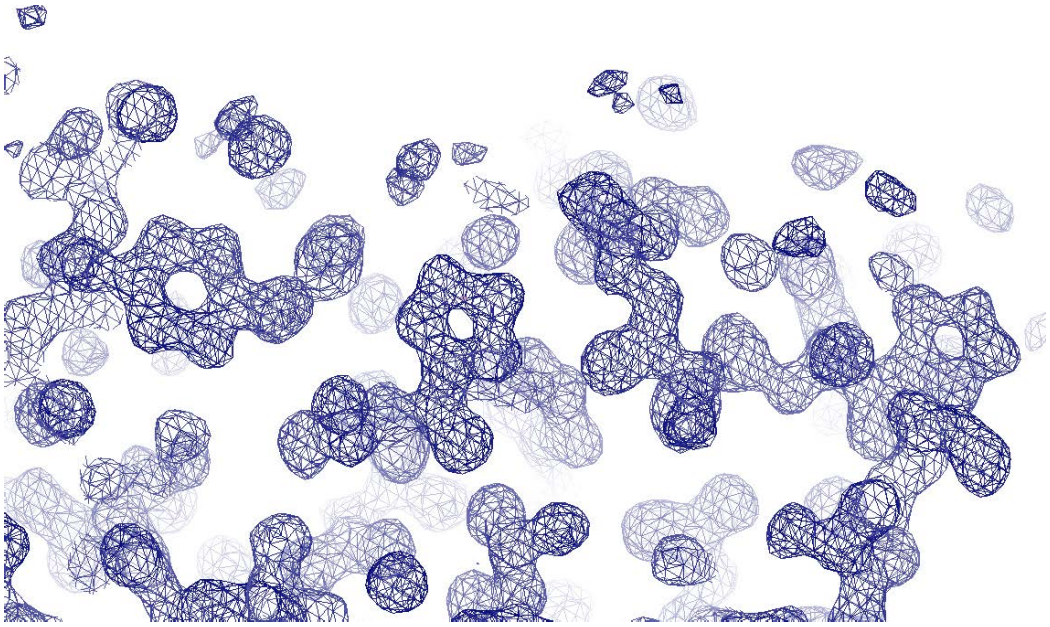


Figure 4.37 – Carte de densité électronique **expérimentale** (contourée à 1,5 σ) améliorée par aplatissement de solvant d'un cristal dérivé de TdTHAU + $[Eu(HETDPA)_3]^{3-}$, phasage MAD, résolution : 1,20 Å

Affinement des structures et description du mode de fixation du $[Eu(HETDPA)_3]^{3-}$ sur le HEWL

Le **tableau 4.18** suivant donne les valeurs des indicateurs statistiques qui permettent de juger la qualité du modèle HEWL+$[Eu(HETDPA)_3]^{3-}$ affiné.

	$[Eu(HETDPA)_3]^{3-}$	
	$HEWL_{co-cristallisation}$	$HEWL_{trempage}$
Rapport $\frac{Qté\,de\,complexe}{Qté\,de\,protéine}t_0$	3,0	$\approx$ 10,8
Code RCSB-PDB	4BAF	-
Résolution (Å)	34,78 - 1,51	39,10 - 1,93
R_{work} (%)	15,64	18,98
R_{free} (%)	17,87	23,19
Nbre de réflexions utilisées	18847	9233
Composition (nbre d'atomes)		
Protéine	1048	1048
H2O	148	152
Ions	12	13
Ligands	62	12
Res. hors Ramachandran (%)	0	0
Ecarts types (globaux)		
$Liaisons_{Longueurs}$ (Å)	0,011	0,003
$Liaisons_{Angles}$ (°)	1,384	0,718
B_{moyen} protéine (Å^2)	15,87	16,90
B_{min} protéine (Å^2)	7,97	8,25
B_{max} protéine (Å^2)	47,11	38,02
B_{moyen} eau (Å^2)	30,00	29,27
B_{moyen} ions lanthanide (Å^2)	10,72	19,57
B_{moyen} ligands (Å^2)	14,10	21,42
Site 1_{Occ} (%)	45	19
Site 2_{Occ} (%)	-	7

Tableau 4.17 – Résumé des statistiques d'affinement pour la structure du HEWL avec le $[Ln(HETDPA)_3]^{3-}$

Au regard de la structure du HEWL modélisée avec le complexe $[Eu(HETDPA)_3]^{3-}$, on constate que son mode de fixation n'est pas différent de celui du complexe $[Eu(HMTDPA)_3]^{3-}$ précédemment étudié. En ce sens, la fonction hydroxyethyle, seul changement au niveau de la structure, n'apporte quasiment rien par rapport à la fonction hydroxymethyl.

Dans les cristaux dérivés de lysozyme, le $[Eu(HETDPA)_3]^{3-}$ occupe donc les mêmes sites de fixation que le $[Eu(HMTDPA)_3]^{3-}$. En co-cristallisation, le premier site est peuplé à 45 % alors qu'à l'emplacement du second site, l'absence de densité ne permet pas de placer le diffuseur anomal. En trempage, le premier site n'est plus occupé qu'à 19 % alors que le second l'est à 7 %.

Toutes les observations concernant le mode du fixation du complexe $[Ln(HETDPA)_3]^{3-}$ à la surface du lysozyme de blanc d'œuf de poule ont été réalisées à partir du cristal dérivé obtenu par co-cristallisation.

Site 1, description de l'interaction du premier ligand avec la protéine De la même façon que pour le $[Ln(HMTDPA)_3]^{3-}$, le premier ligand du complexe $[Ln(HETDPA)_3]^{3-}$ établit une liaison hydrophobe CH-π entre son cycle triazole et le cycle benzénique du noyau indole du Trp123. La liaison hydrogène entre l'un de ses groupements carboxylates et le groupement $N\zeta H_2$ de la Lys33 est toujours présente. Ces deux interactions se font avec la molécule de référence (en vert). La seule différence avec

le ligand HMTDPA réside au niveau de la troisième interaction : une liaison polaire s'établit entre la fonction hydroxyethyl du ligand HETDPA et la fonction carboxyle de l'acide aminé histidie 117 d'une molécule symétrique (en orange) **[Figure 4.38]**.

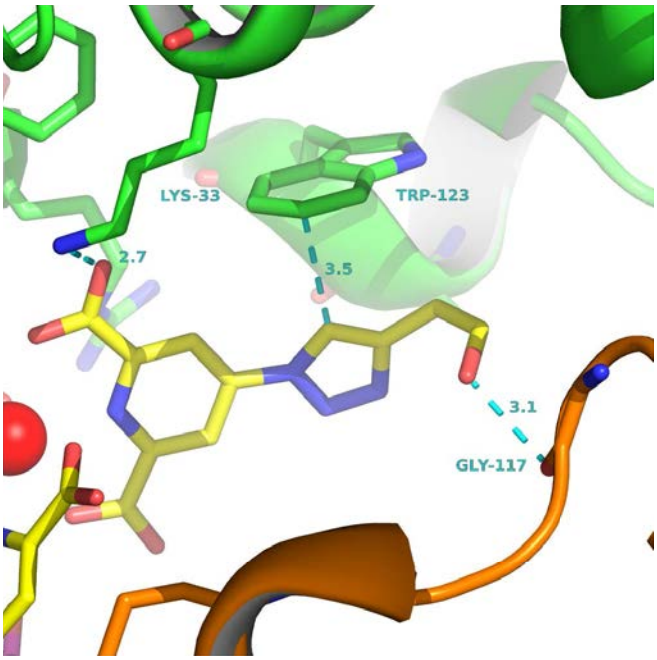


Figure 4.38 – Mode de fixation du complexe $[Ln(HETDPA)_3]^{3-}$ à la surface du HEWL. Cristal dérivé obtenu par co-cristallisation. 1^{er}ligand.

Site 1, description de l'interaction du second ligand avec la protéine On retrouve l'empilement-π entre le cycle triazole du second ligand et le cycle benzénique du noyau indole du Trp62 de la molécule symétrique (en rose). Je n'ai pas trouvé d'interaction mettant en cause la queue hydroxyethyle pour ce ligand **[Figure 4.39]**.

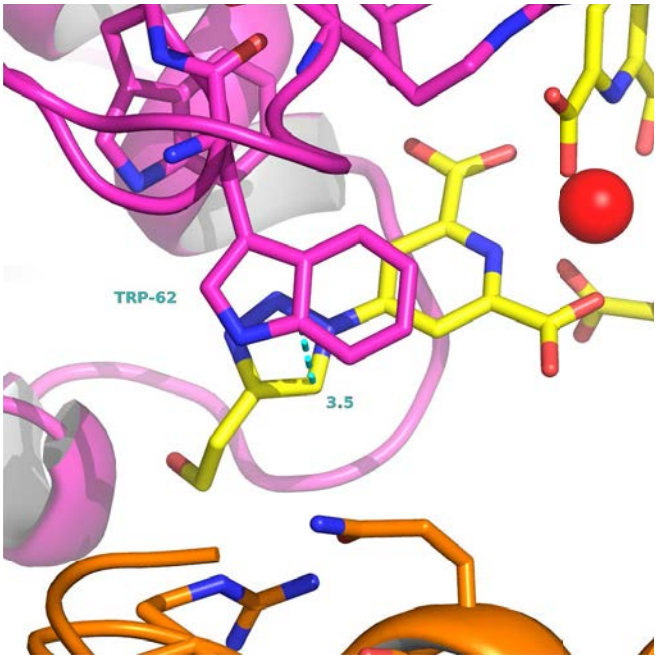


Figure 4.39 – Mode de fixation du complexe $[Ln(HETDPA)_3]^{3-}$ à la surface du HEWL. Cristal dérivé obtenu par co-cristallisation. 2^d ligand.

Site 1, description de l'interaction du troisième ligand avec la protéine La fixation du troisième ligand HETDPA est la même que pour le ligand HMTDPA. On modélise clairement une liaison hydrogène entre un de ses groupements carboxylates et le groupement $N_\varepsilon H$ de l'Arg5. Malgré la distance mesurée de 3,7 Å, on peut signaler une liaison hydrogène supplémentaire entre une des fonctions carboxylates du complexe et le groupement NH1 de l'Arg73 de la molécule symétrique (en rose) **[Figure 4.40]**.

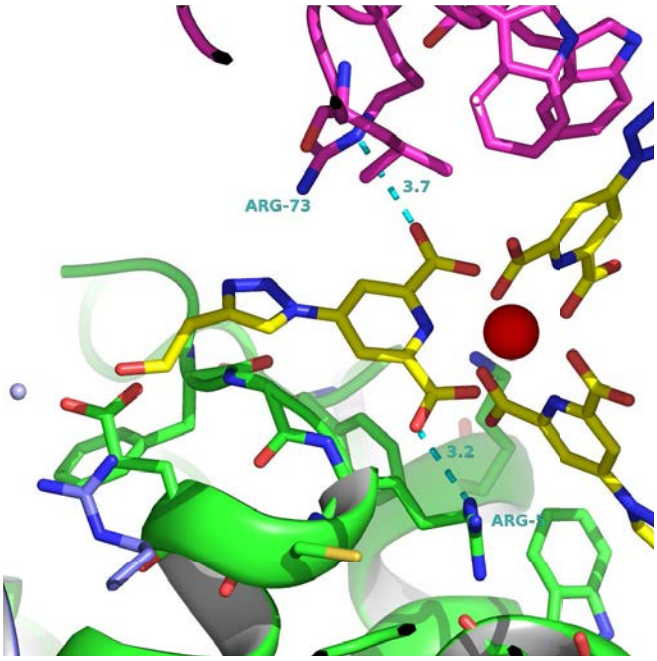


Figure 4.40 – Mode de fixation du complexe $[Ln(HETDPA)_3]^{3-}$ à la surface du HEWL. Cristal dérivé obtenu par co-cristallisation. $3^{ème}$ ligand.

On constate donc que le mode de fixation du complexe $[Ln(HETDPA)_3]^{3-}$ ne diffère de celui du $[Ln(HMTDPA)_3]^{3-}$ que par une seule liaison polaire supplémentaire. La fonction hydroxyethyle du ligand HETDPA, plus longue que la fonction hydroxymethyl du HMTDPA permet d'atteindre la molécule symétrique pour y créer cette liaison supplémentaire.

Cependant, nous n'avons pas pu mettre en évidence le mode de fixation du complexe dans le deuxième site Trp62, celui-ci étant inoccupé. Ceci peut s'expliquer par la faible concentration à laquelle le complexe $[Eu(HETDPA)_3]^{3-}$ a été utilisé pour l'expérience de co-cristallisation. Bien qu'il y ait trois fois plus de molécules de complexe que de molécules de protéine dans la goutte, cette proportion n'est visiblement pas suffisante pour peupler le second site. D'après le mécanisme de cristallisation des cristaux dérivés précédemment proposé **[Figure 4.29 page 142]**, les molécules de complexe se fixent en priorité aux "sites tryptophane".

Or, à si basse concentration, il est possible qu'il n'y ait pas assez de complexes pour saturer tous les sites à la surface du lysozyme. Comme le site Trp123 est celui qui propose à la fois les interactions hydrophobes CH-π et π-π, c'est le site qui est peuplé de façon préférentielle.

Cette expérience de co-cristallisation menée avec le complexe $[Ln(HETDPA)_3]^{3-}$ sur le HEWL confirme qu'*a priori*, les modes de fixation préférés des complexes $[Ln(TDPA)_3]^{x}$ sont bien les interactions hydrophobes CH-π et π-π entre leur noyau triazole et le noyau indole des acides aminés tryptophane. De plus, cette expérience a permis de confirmer le mécanisme de cristallisation que nous avons précédemment proposé et a confirmé le réalisme de celui-ci.

Affinement des structures et description du mode de fixation du $[Eu(HETDPA)_3]^{3-}$ sur TdTHAU

Nous avons pu obtenir un modèle affiné du complexe $[Eu(HETDPA)_3]^{3-}$ fixé à la surface de la protéine TdTHAU. Les indicateurs statistiques relatifs à l'affinement de ce modèle sont indiqués dans le **tableau 4.18**.

	$[Eu(HETDPA)_3]^{3-}$
	$TdTHAU_{co-cristallisation}$
Rapport $\frac{Qté\,de\,complexe}{Qté\,de\,protéine}_{t_0}$	2,3
Code RCSB-PDB	4BAR
Résolution (Å)	31,50 - 1,30
R_{work} (%)	13,70
R_{free} (%)	15,55
Nbre de réflexions utilisées	78332
Composition (nbre d'atomes)	
Protéine	1592
H2O	411
Ions	1
Ligands	60
Res. hors Ramachandran (%)	0
Ecarts types (globaux)	
$Liaisons_{Longueurs}$ (Å)	0,013
$Liaisons_{Angles}$ (°)	1,613
B_{moyen} protéine (Å^2)	11,77
B_{min} protéine (Å^2)	5,42
B_{max} protéine (Å^2)	32,02
B_{moyen} eau (Å^2)	28,29
B_{moyen} ions lanthanide (Å^2)	17,77
B_{moyen} ligands (Å^2)	20,51
$Site_{Occ}$ (%)	49

Tableau 4.18 – Résumé des statistiques d'affinement de la structure du TdTHAU avec le $[Ln(HETDPA)_3]^{3-}$

A la surface de la thaumatine de *Thaumatococcus daniellii*, le $[Eu(HETDPA)_3]^{3-}$ se place dans la même poche que le $[Ln(HMTDPA)_3]^{3-}$, les ligands dans la même orientation.

Bien que le centre du complexe soit légèrement décalé par rapport à celui du complexe $[Ln(HMTDPA)_3]^{3-}$, leurs modes de fixation se sont avérés identiques. Résoudre cette structure a donc permis de confirmer la présence de la liaison hydrogène entre le groupement NH_2 d'une des deux conformations de l'Arg79 et le groupement carboxylate du premier et du second ligand. Les fonctions hydroxyethyles des ligands ne permettent pas à ce complexe d'établir de nouvelles liaisons avec la surface de la protéine : ces groupements sont toujours orientés vers le solvant, au cœur de la poche **[Figure 4.41]**.

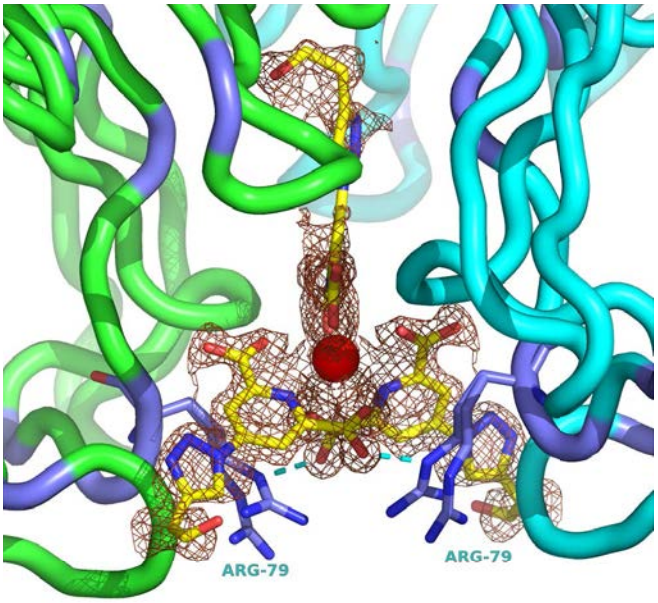


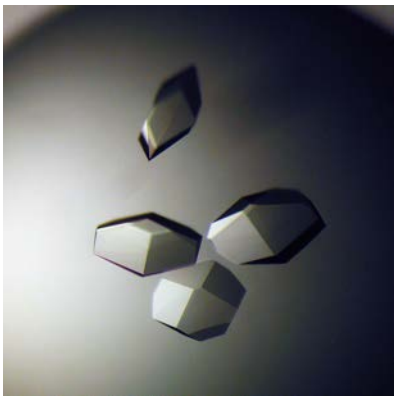
Figure 4.41 – Carte de densité électronique du $[Eu(HETDPA)_3]^{3-}$ après affinement, à la surface de la TdTHAU.

4.6.1.2 Récapitulatif des essais réalisés avec le complexe $Eu(HECTDPA)_3$

Le complexe $Eu(HECTDPA)_3$ se distingue des autres complexes de la famille des triazoledipicolinates par sa charge globale nulle et son encombrement. Ceci est du à la fonction éthylcholine **[Figure 4.34 page 147]**. De plus, sa taille correspond à une sphère de 28 Å environ (alors que celle des complexes $[Eu(HMTDPA)_3]^{3-}$ et $[Eu(HETDPA)_3]^{3-}$ est de 22 Å et 24 Å, respectivement). Nos observations ont révélé que, tant pour l'obtention de cristaux dérivés que pour son mode de fixation, le complexe $Eu(HECTDPA)_3$ diffère des deux autres complexes : $[Eu(HMTDPA)_3]^{3-}$ et $[Eu(HETDPA)_3]^{3-}$.

Obtention de cristaux dérivés de HEWL avec le $Eu(HECTDPA)_3$

Contrairement aux complexes $[Eu(HMTDPA)_3]^{3-}$ et le $[Eu(HETDPA)_3]^{3-}$, je n'ai jamais constaté de double faciès cristallin dans les gouttes de co-cristallisation. Qu'elle que soit la concentration en complexe utilisée, les cristaux dérivés obtenus présentent des faces unies et ressemblent aux cristaux natifs de lysozyme mais aplatis dans une direction **[Figure 4.42]**. Pour du lysozyme et du complexe mélangés dans les mêmes rapports de concentration que pour les $[Eu(HMTDPA)_3]^{3-}$ et le $[Eu(HETDPA)_3]^{3-}$, j'ai toujours observé, avec le $Eu(HECTDPA)_3$ des cristaux dérivés bien plus gros et moins nombreux dans la goutte de cristallisation.



(a) Cristaux dérivés de HEWL (10 $mg.mL^{-1}{}_{t_0}$) avec 10 mM_{t_0} $Eu(HMTDPA)_3$], lumière visible

(b) Cristaux dérivés de HEWL (10 $mg.mL^{-1}{}_{t_0}$) avec 10 mM_{t_0} $Eu(HMTDPA)_3$], UV proches

Figure 4.42 – Cristaux dérivés de HEWL par co-cristallisation avec le complexe $Eu(HECTDPA)_3$

Intégration des données de diffraction issues de cristaux dérivés de HEWL avec le $Eu(HECTDPA)_3$]

Plusieurs jeux de données à haute résolution ont pu être enregistrés à partir des cristaux dérivés obtenus par co-cristallisation, données de qualité qui ont conduit à la résolution de la structure du complexe et de la protéine **[Tableaux 4.19 et 4.20]**.

	$Eu(HECTDPA)_3$		
	HEWL		
[Complexe]$_{t_0}$ (mM)	2,7		10
[Protéine]$_{t_0}$ (mM)	0,46		0,67
Rapport $\frac{Qté\,de\,complexe}{Qté\,de\,protéine}{}_{t_0}$	5,87		14,92
Méthode	co-cristallisation		
Source synchrotron	ESRF		
Ligne de lumière	FIP-BM30A		
MAD	oui		non
λ (Å)	1,776 ($Eu_{L_{III}}$)	0,979	1,776 ($Eu_{L_{III}}$)
Groupe d'espace	$P4_32_12$		$P4_32_12$
Paramètres de maille (Å)	a = 77,027, c = 38,110		a = 77,340, c = 38,080
Contenu unité asymétrique	1 molécule		
Résolution (Å)	38,51 - 1,93 (2,04 - 1,93)	38,15 - 1,36 (1,43 - 1,36)	38,69 - 1,94 (2,04 - 1,94)
Nombre de réflexions uniques	9018 (1251)	25390 (3604)	9105 (1263)
R_{merge} (%)	4,9 (20,9)	5,6 (48,9)	6,7 (28,5)
R_{pim} (%)	2,0 (8,9)	2,9 (26,0)	3,4 (15,1)
Sig_{ano} XDS	4,84 (1,75)	1,17 (0,76)	2,91 (1,11)
R_{ano} (%)	9,2 (13,3)	3,3 (16,0)	9,4 (15,8)
$\frac{I}{\sigma(I)}$	12,2 (3,6)	8,3 (1,5)	8,7 (2,6)
Complétude (%)	99,7 (97,7)	99,8 (99,0)	99,6 (97,6)
Redondance	12,8 (11,8)	8,5 (8,4)	8,5 (7,9)

Tableau 4.19 – Statistiques d'intégration des jeux de données de diffraction de cristaux dérivés de HEWL + $Eu(HECTDPA)_3$

	Eu(HECTDPA)$_3$	
	HEWL	
[Complexe**]$_{t_0}$ (mM)**	12	
[Protéine**]$_{t_0}$ (mM)**	0,67	
Rapport $\frac{Qté\ de\ complexe}{Qté\ de\ protéine}\ t_0$	17,91	
Méthode	co-cristallisation	
Source synchrotron	ESRF	
Ligne de lumière	FIP-BM30A	
MAD	oui	
λ (Å)	1,776 (Eu$_{L_{III}}$)	0,979
Groupe d'espace	$P4_32_12$	
Paramètres de maille (Å)	a = 77,240, c = 38,362	
Contenu unité asymétrique	1 molécule	
Résolution (Å)	38,62 - 1,93 (2,03 - 1,93)	38,75 - 1,21 (1,27 - 1,21)
Nombre de réflexions uniques	9212 (1291)	36469 (5190)
R$_{merge}$ (%)	5,3 (20,8)	5,1 (59,8)
R$_{pim}$ (%)	3,3 (13,4)	6,5 (34,4)
Sig$_{ano}$ XDS	3,09 (1,32)	1,19 (0,80)
R$_{ano}$ (%)	9,5 (15,9)	3,5 (27,3)
$\frac{I}{\sigma(I)}$	10,7 (3,7)	10,8 (1,3)
Complétude (%)	99,7 (98,2)	99,7 (98,8)
Redondance	6,5 (6,3)	7,6 (7,2)

Tableau 4.20 – Statistiques d'intégration des jeux de données de diffraction de cristaux dérivés de HEWL + Eu(HECTDPA)$_3$, suite.

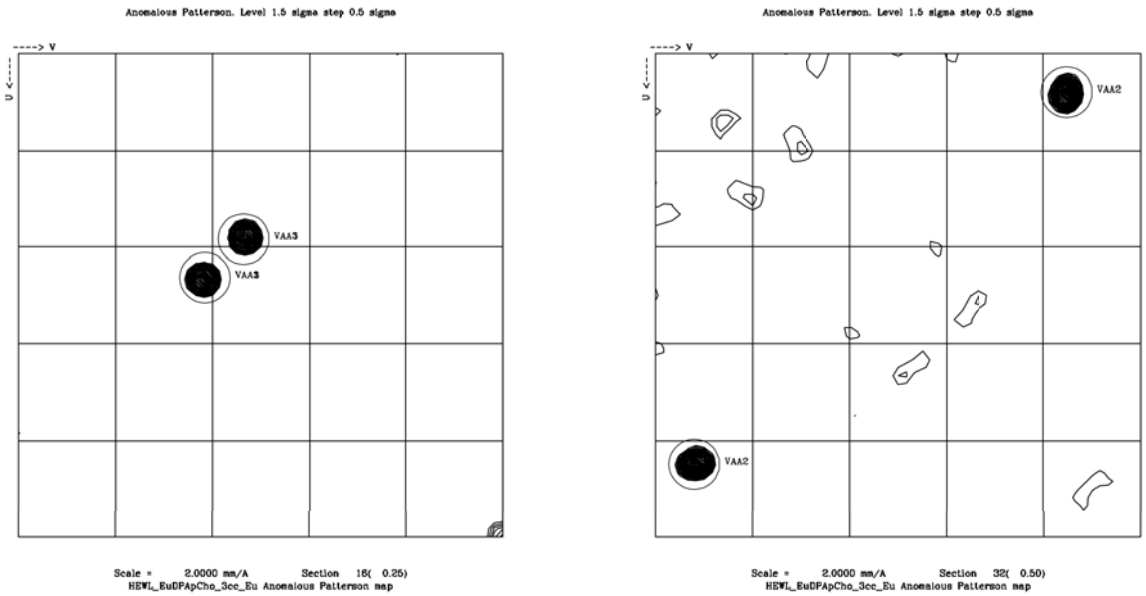
Phasage expérimental avec le Eu(HECTDPA)$_3$

Les jeux de données de diffraction dont les statistiques sont décrites ci-dessus ont tous conduit à un phasage d'excellente qualité **[Figure 4.21]**. De plus, la recherche de pics après calcul de la Fourier anomale montre que deux autres jeux de données SAD, enregistrés hors seuil de l'europium, ont aussi conduit à la mesure de fortes différences anomales. Ces deux jeux de données sont issus de l'expérience de diffraction de cristaux apparus suite à la co-cristallisation de la protéine HEWL concentrée à 0,46 mM$_{t_0}$ avec, respectivement, 2,7 mM$_{t_0}$ et 12 mM$_{t_0}$ de Eu(HECTDPA)$_3$.

	Eu(HECTDPA)$_3$		
	$HEWL_{co-cristallisation}$		
Rapport $\frac{Qté\,de\,complexe}{Qté\,de\,protéine}t_0$	5,87	14,92	17,91
Type de données	MAD	$SADEu_{L_{III}}$	MAD
Résolution (Å)	38,15 - 1,36	38,69 - 1,94	38,75 - 1,21
Pics$_{Fa}$ (Pic$_{Soufre}$) pk (σ)	59,24 (6,20)	42,77 (6,45)	60,30 (5,50)
Pics$_{Fa}$ (Pic$_{Soufre}$) rm (σ)	56,56 (8,04)	-	72,22 (7,86)
$\mathbf{FOM}_{sharp}^{acen.}$ / $\mathbf{FOM}_{sharp}^{cen.}$	0,47 / 0,43	0,52 / 0,12	0,38 / 0,14
Nombre de sites	1	1	1
Résolution $\mathbf{FOM}_{sharp}^{acen.} \approx$ 0,5 (Å)	1,88	2,10	1,68
$\varphi\mathbf{p}_{sharp\,acen.}^{\neq anomales}$ / $\varphi\mathbf{p}_{sharp\,acen.}^{\neq isomorphes}$	1,33 / 0,00	5,28 / 0,00	1,32 / 0,00
$\mathbf{SIGMAA}_{solomon}$	0,55	0,50	0,56
Arp/Warp (construit / total)	50 / 129	82 / 129	104 / 129

Tableau 4.21 – Statistiques de phasage expérimental pour des cristaux dérivés de HEWL avec le complexe Eu(HECTDPA)$_3$.

Le pouvoir phasant du complexe Eu(HECTDPA)$_3$ a aussi clairement été observable lors de l'examen de carte de Patterson anomale **[Figure 4.43]** et de la carte de densité électronique expérimentale améliorée par modification de densité. La figure 4.44 montre que l'empilement π entre un des ligands du complexe et le tryptophane du lysozyme est déjà visible dans la carte de densité électronique expérimentale.


Anomalous Patterson. Level 1.5 sigma step 0.5 sigma
VAA3
VAA3
Scale = 2.0000 mm/A Section 16(0.25)
HEWL_EuDPApCho_3cc_Eu Anomalous Patterson map
Anomalous Patterson. Level 1.5 sigma step 0.5 sigma
VAA2
VAA2
Scale = 2.0000 mm/A Section 32(0.50)
HEWL_EuDPApCho_3cc_Eu Anomalous Patterson map

(a) Carte de Patterson anomale, section de Harker w = $\frac{1}{4}$

(b) Carte de Patterson anomale, section de Harker w = $\frac{1}{2}$

Figure 4.43 – Deux sections de Harker de la carte de Patterson anomale du cristal dérivé de HEWL par co-cristallisation avec 2,7 mM_{t_0} du Eu(HECTDPA)$_3$ (jeu de donnée MAD)

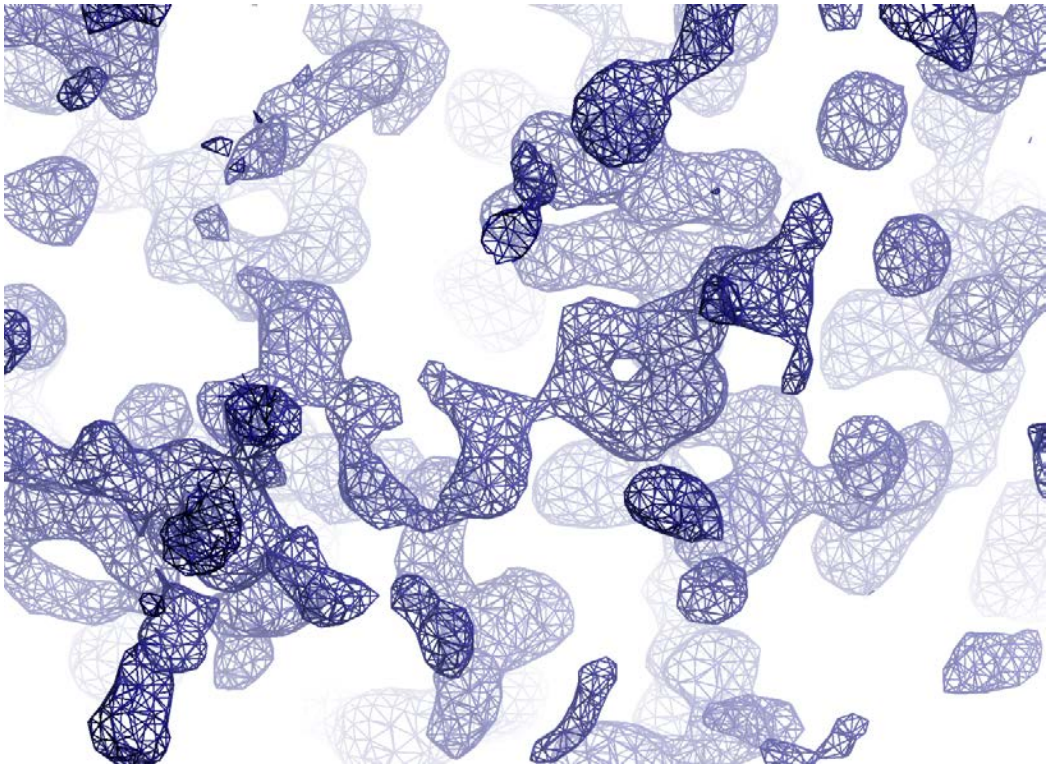


Figure 4.44 – Carte de densité électronique **expérimentale** améliorée par aplatissement de solvant d'un cristal dérivé de HEWL + $Eu(HECTDPA)_3$, contourée à 1,0 σ, phasage MAD, 1,21 Å de résolution

Affinement des structures et description du mode de fixation du $Eu(HECTDPA)_3$

Trois structures provenant de cristaux dérivés avec trois concentrations différentes de complexe ont pu être résolues à haute résolution. Cela nous a permis encore une fois de regarder l'influence de la concentration en complexe sur les taux d'occupation des sites de fixation.

Que le complexe $Eu(HECTDPA)_3$ soit ajouté à 2,7 ou 12 mM_{t_0}, la fixation à la protéine se fait majoritairement dans le site 1 (Trp123). Les taux d'occupation affinés pour ce site plafonnent à 62 % dans les structures issues d'un cristal dérivé avec 2,7 mM_{t_o} de complexe **[Tableau 4.22]**.

	Eu(HECTDPA)$_3$		
	$HEWL_{co-cristallisation}$	$HEWL_{co-cristallisation}$	$HEWL_{co-cristallisation}$
Rapport $\frac{Qté\ de\ complexe}{Qté\ de\ protéine}t_0$	5,87	14,92	17,91
Code RCSB-PDB	-	-	4BAP
Résolution (Å)	34,47 - 1,36	38,69 - 1,94	38,75 - 1,21
R_{work} (%)	13,86	17,75	15,81
R_{free} (%)	17,29	20,22	17,83
Nbre de réflexions utilisées	25342	16468	68533
Composition (nbre d'atomes)			
Protéine	1113	1105	1100
H2O	171	111	146
Ions	11	12	11
Ligands	100	76	4
Res. hors Ramachandran (%)	0	0	0
Ecarts types (globaux)			
Liaisons$_{Longueurs}$ (Å)	0,018	0,005	0,008
Liaisons$_{Angles}$ (°)	0,278	0,900	1,253
B_{moyen} protéine (Å^2)	14,70	15,54	13,35
B_{min} protéine (Å^2)	5,9	4,9	6,50
B_{max} protéine (Å^2)	37,35	54,06	42,98
B_{moyen} eau (Å^2)	28,80	27,26	27,40
B_{moyen} ions lanthanide (Å^2)	15,62	19,89	11.63
B_{moyen} ligands (Å^2)	23,04	15,30	17,00
Site 1 $_{Occ}$ (%)	62	48	62
Site 2 $_{Occ}$ (%)	-	4	-

Tableau 4.22 – Résumé des statistiques d'affinement de la structure du HEWL avec le $Ln(HECTDPA)_3$

Description du mode de fixation du Eu(HECTDPA)$_3$ à la surface de HEWL

Description des interactions impliquées par le 1er ligand Avec la molécule de référence (en vert), le premier ligand ne crée plus la liaison hydrophobe CH-π entre son cycle triazole et le cycle benzénique du noyau indole du Trp123 mais une liaison π-π comme le second ligand le fait avec le Trp62. La liaison hydrogène entre l'un de ses groupements carboxylates et le groupement $N\zeta H_2$ de la Lys33 est toujours présente, cette lysine 33 adoptant une seconde conformation. Une autre différence, plus légère, réside dans une interaction polaire entre sa fonction hydroxyethyle et la fonction carboxyle de la Thr118 d'une molécule symétrique (en rose) via un pont impliquant une molécule d'eau, interaction permise grâce à la longueur de la fonction hydroxyethylcholine **[Figure 4.45]**.

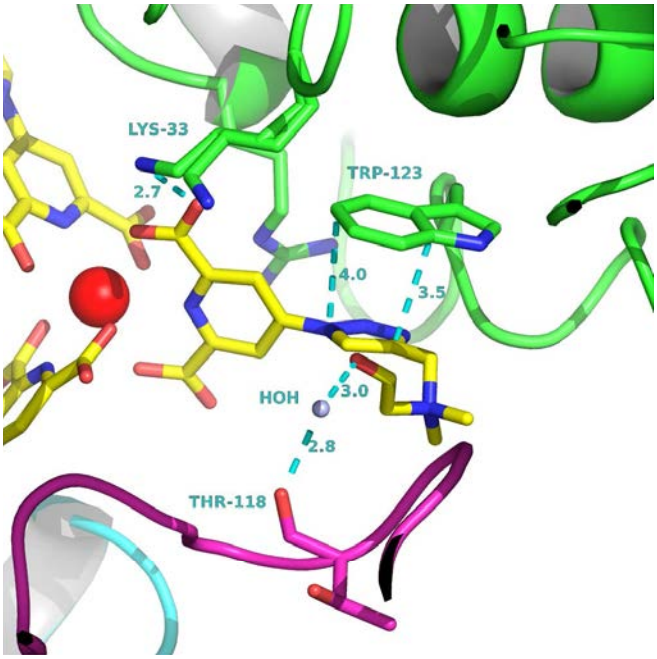


Figure 4.45 – Mode de fixation du complexe Ln(HECTDPA)$_3$ à la surface du HEWL. Cristal dérivé par co-cristallisation. 1^{er} ligand.

Description des interactions impliquées par le 2^d ligand Le second ligand établit toujours l'empilement-π entre son cycle triazole et le cycle benzénique du noyau indole du Trp62 d'une molécule symétrique (en bleu) **[Figure 4.46]**. En outre, la longueur de sa "queue choline" lui permet d'établir encore une fois une liaison polaire impliquant une molécule d'eau avec la fonction carboxyle de l'Ala107 de cette même molécule symétrique.

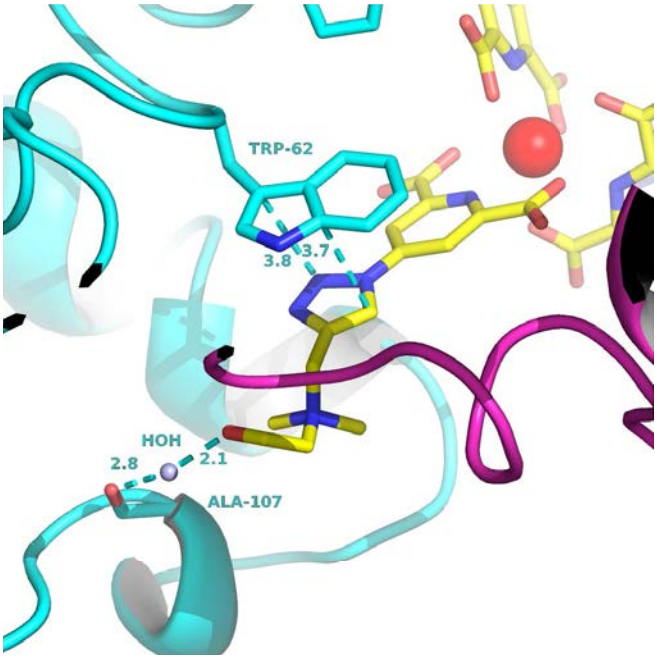


Figure 4.46 – Mode de fixation du complexe Ln(HECTDPA)$_3$ à la surface du HEWL. Cristal dérivé par co-cristallisation. 2^{nd} ligand.

Description des interactions impliquées par le 3ème ligand A l'instar des ligands HMTDPA et HETDPA, le mode de fixation du troisième ligand HECTDPA à la protéine comprend toujours la liaison hydrogène entre un de ses groupements carboxylates et le groupement $N_\varepsilon H$ de l'Arg5. Et cette fois-ci, nous pouvons confirmer l'autre liaison hydrogène entre son autre fonction carboxylate et le groupement NH1 de l'Arg73 de la molécule symétrique (en bleu). Une liaison polaire supplémentaire semble se créer entre la fonction hydroxyle de la "queue choline" et l'acide aminé Glu7 de la molécule de référence **[Figure 4.47]**.

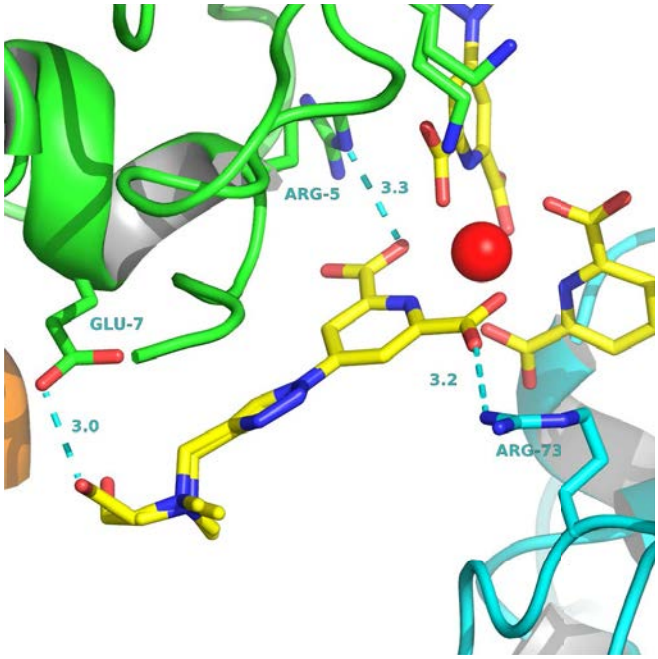


Figure 4.47 – Mode de fixation du complexe $Ln(HECTDPA)_3$ à la surface du HEWL. Cristal dérivé par co-cristallisation. 3ème ligand.

Si nous résumons l'ensemble des observations recueillies lors de cette étude concernant le complexe $Ln(HECTDPA)_3$, on constate qu'il se distingue des complexes $[Ln(HMTDPA)_3]^{3-}$ et $[Ln(HETDPA)_3]^{3-}$ sur un certain nombre de points :

- ▷ Que le complexe ait été introduit dans la goutte à une concentration de 2,7 mM_{t_0} ou 10 mM_{t_0}, aucun cristal au faciès torturé n'est apparu dans la goutte.
- ▷ Aucun cristal dérivé n'a pu être obtenu par trempage. Non pas que les cristaux natifs soient détériorés par le procédé de trempage, mais suite à l'analyse des données de diffraction, nous n'avons jamais constaté de fixation du complexe dans les cristaux dérivés par cette méthode. Sur l'ensemble des jeux de données enregistrés à partir de cristaux "trempés", je n'ai vu qu'une seule fois un très faible pic à 11,14 σ en Fourier anomale, calculé pour un cristal de lysozyme trempé pendant 10 minutes dans une solution contenant 2,7 mM de $Eu(HECTDPA)_3$.
- ▷ Seuls des cristaux dérivés de lysozyme de blanc d'œuf de poule ont pu être obtenus par co-cristallisation. Le complexe ne s'est jamais fixé sur la thaumatine.
- ▷ Dans les cristaux dérivés de lysozyme, seul le site Trp123 est occupé par le complexe.

▷ La liaison CH-π entre le noyau triazole d'un ligand et le tryptophane 123, observée avec les complexes $[Ln(HMTDPA)_3]^{3-}$ et $[Ln(HETDPA)_3]^{3-}$, est systématiquement remplacée par un empilement π.

Compte tenu de nos précédentes observations quant à l'apparition des "mauvais cristaux", l'interaction du complexe $Ln(HECTDPA)_3$ avec les acides aminés arginine doit être beaucoup plus faible qu'avec les complexes $[Ln(HMTDPA)_3]^{3-}$ et $[Ln(HETDPA)_3]^{3-}$. La charge globale du complexe est neutre. Ainsi, seules les interactions avec les acides aminés tryptophane sont prépondérantes. Donc même si ce complexe est ajouté à plus haute concentration, la liaison aux acides aminés arginine n'est pas favorisée. D'une part, ceci explique pourquoi je n'ai jamais observé de mauvais cristaux, contrairement aux complexes $[Ln(HMTDPA)_3]^{3-}$ et $[Ln(HETDPA)_3]^{3-}$ et, d'autre part, cela explique aussi pourquoi ce complexe ne se lie pas à la thaumatine. Cette protéine ne propose en effet que des acides aminés arginine pour établir des liaisons hydrogènes et les interactions hydrophobes étant le mode de fixation exclusif du complexe $Ln(HECTDPA)_3$, il ne se fixe pas à sa surface.

Tout comme nous l'avions mentionné pour le complexe $[Ln(HMTDPA)_3]^{3-}$, le fait d'avoir plusieurs structures résolues avec le même complexe permet d'avoir une idée du positionnement statistique des ligands du complexe. La **figure 4.48** montre que, comme le laissait présupposer les facteurs d'agitation thermique et la double conformation modélisée dans la structure, la "queue" hydroxyethylcholine est assujétie à un mouvement rotationnel autour de l'axe que forme la liaison entre le N^+ et le carbone qui suit le noyau triazole. Ceci est valable pour les trois ligands qui composent le complexe. Cette observation confirme que la "queue choline" ne joue pas un rôle direct dans la fixation du complexe : elle n'est pas impliquée dans des liaisons fortes avec la protéine. Mais elle joue un rôle indirect : porteuse de la charge qui rend le complexe neutre, elle a conféré à l'ensemble du complexe de nouvelles propriétés de fixation.

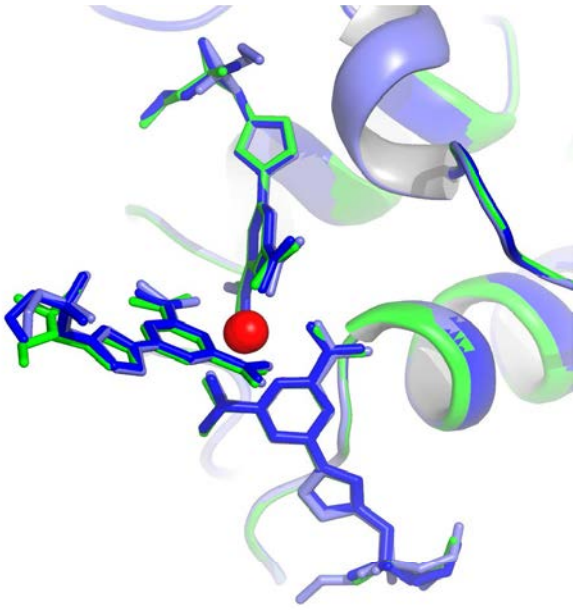


Figure 4.48 – Vue du site de fixation 1 du $Eu(HECTDPA)_3$ après superposition de 3 structures.

Pour l'instant, nous ne pouvons expliquer pourquoi l'interaction CH-π est systématiquement remplacée par un empilement π. Nous avions envisagé que l'encombrement stérique induit par la présence de la "queue choline" pouvait contraindre la rotation du noyau triazole. Mais la **figure 4.48** semble prouver le contraire. En effet, la position des noyaux triazoles semble toujours conservée alors que la fonction ethylcholine est assujettie à une agitation.

Le complexe $Ln(HECTDPA)_3$ constitue donc une alternative de choix pour l'insertion de diffuseurs anomaux dans les protéines. Appartenant à la famille des tris-triazoledipicolinates, il conserve la très forte affinité pour les surfaces protéiques mais s'affranchit de l'effet supramoléculaire dû à la fixation aux acides aminés arginines qui, dans le cas de cette famille de complexe, conduit aux mauvais cristaux.

4.6.1.3 Récapitulatif des essais réalisés avec le complexe $[Eu(CTDPA)_3]^{6-}$

Cristallisation, enregistrement de données et tentative de phasage

Le complexe $[Eu(CTDPA)_3]^{6-}$ est le dernier membre de la famille des tris-triazoledipicolinates que nous présentons dans cette étude. Préparé selon le même protocole que celui du complexe $[Eu(HETDPA)_3]^{3-}$, nous n'avons jamais pu mettre aussi clairement en évidence la fixation de ce nouveau complexe $[Eu(CTDPA)_3]^{6-}$ dans un cristal de protéine et ce, même avec le lysozyme et la thaumatine.

L'ajout d'une solution de $[Eu(CTDPA)_3]^{6-}$ concentrée à 4 mM n'a pas empêché la croissance de cristaux de lysozyme, de thaumatine ou encore de HvMalDH. Mais ce complexe n'a jamais permis de mener à terme un phasage expérimental avec ces protéines. Même après remplacement moléculaire, l'absence de densité électronique pour le ligand ne permet pas de déterminer son propre mode de fixation. Dans le cas de l'enzyme HvMalDH, comme avec les autres complexes de la même famille, aucune fixation n'a pu être décelée. En ce qui concerne le lysozyme et la thaumatine, le constat est moins tranché : le calcul de synthèses de Fourier anomales a permis de révéler que ce complexe se fixe.

Dans le cas du lysozyme, seul le site 1 (Trp123) s'avère en partie peuplé par le $[Eu(CTDPA)_3]^{6-}$. Un premier jeu de données, enregistré au seuil d'absorption L_{III} de l'europium à partir d'un cristal obtenu par co-cristallisation en présence de 1,3 mM de complexe, a conduit au calcul d'un pic à 13,0 σ par synthèse de Fourier anomale. Un second jeu de données enregistré au seuil d'absorption L_{III} de l'europium, mais à partir d'un cristal dérivé obtenu par trempage dans une liqueur contenant 2,7 mM de complexe, a conduit à la présence d'un pic de 22,0 σ.

Cependant, au cours des essais de co-cristallisation du lysozyme, le complexe $[Eu(CTDPA)_3]^{6-}$ a conduit à l'obtention d'un très grand nombre de "mauvais cristaux". Seuls quelques cristaux, co-cristallisés avec 1,3 mM de complexe, ont conduit à des profils de diffraction enregistrables. Dès que la concentration en complexe est supérieure à 3 mM, la co-cristallisation conduit à des cristaux au faciès torturé présentant un mauvais profil de diffraction **[Figure 4.49]**.

Figure 4.49 – Cristal de HEWL co-cristallisé en présence de 2,7 mM de $[Eu(CTDPA)_3]^{3-}$

En ce qui concerne les essais avec la thaumatine, un jeu de données SAD, enregistré au seuil d'absorption L_{III} de l'europium, à partir d'un cristal co-cristallisé avec 1,3 mM de complexe, à conduit à la présence d'un pic de 38,5 σ par synthèse de Fourier anomale. L'intensité de ce pic est cependant à relativiser car, comme pour le complexe $[Ln(HMTDPA)_3]^{3-}$, le lanthanide se situe sur l'axe de symétrie d'ordre 2.

Le $[Ln(CTDPA)_3]^{6-}$ comme illustration du modèle proposé pour expliquer l'effet supramoléculaire

Comme nous l'avons précédemment mentionné, les interactions hydrophobes avec la protéine entre les noyaux triazoles des ligands et les acides aminés tryptophane de la protéine sont le mode de fixation privilégié par les membres de la famille des $[Ln(TDPA)_3]^x$. De par le groupement carboxylate supplémentaire sur chacun de ses ligands, le complexe $[Ln(CTDPA)_3]^{6-}$ présente une charge globale six fois négative. Or il s'agit du complexe qui, une fois introduit dans la goutte, conduit systématiquement au plus grand nombre de cristaux au faciès torturé. Cette forte charge fait basculer son mode de fixation qui s'oriente, alors, non plus vers les interactions hydrophobes mais uniquement vers les liaisons hydrogènes avec les acides aminés arginine et lysine.

En ce sens, le complexe $[Ln(CTDPA)_3]^{6-}$ est l'opposé du complexe $Ln(HECTDPA)_3$: alors que ce dernier, de charge globale neutre, privilégie le mode de fixation impliquant les interactions hydrophobes avec les acides aminés tryptophane, le $[Ln(CTDPA)_3]^{6-}$, six fois chargé négativement, privilégie le mode de fixation impliquant les acides aminés arginine et lysine. Son mode de fixation ressemble alors à celui du tris-dipicolinate de lanthanide, induisant des liaisons supramoléculaires, pour créer un nouvel empilement cristallin. Mais la présence d'un petit nombre de complexes fixés aux acides aminés tryptophane et/ou l'encombrement qu'induit le noyau triazole conduit à un empilement mal organisé et donc aux cristaux présentant un faciès torturé.

4.7 Le $[Ln(HMTDPA)_3]^{3-}$, un complexe modulable : la combinatoire

Le fait d'avoir, d'une part, un complexe composé de trois ligands non-liés covalament entre eux et, d'autre part, plusieurs ligands différents rapidement modifiables par "chimie-click", a ouvert une autre piste : celle de "la combinatoire". J'entends par ce terme le fait de créer un complexe, de type tris-

dipicolinate, non plus avec trois ligands identiques mais avec une combinaison de trois ligands choisis selon ses besoins.

Cette stratégie expérimentale peut solliciter la synthèse chimique d'un "tris-hétérodipicolinate de lanthanide" et elle émane du principe qu'il peut y avoir échange de ligands autour de l'ion lanthanide. Au laboratoire de chimie SEESIB, une expérience qui permet de rapidement mettre en évidence ces échanges a été réalisée. Elle a simplement consisté à ajouter une solution concentrée de sel de terbium dans une solution concentrée en complexe $[Eu(HMTDPA)_3]^{3-}$. La luminescence de la solution qui était alors rouge sous UV est devenue progressivement verte, prouvant qu'il y a eu échange d'ions lanthanide au sein de complexes déjà formés. Nous avons donc aussi envisagé de créer un complexe combiné en complexant nous-même un ion lanthanide avec les ligands de type picolinate présentés dans la sous-section précédente.

Par le terme "combinatoire", on peut aussi sous-entendre l'utilisation simultanée de complexes de lanthanide différents pour augmenter la probabilité de fixation de diffuseurs anomaux sur la protéine. Nous avons mené une expérience de co-cristallisation mettant en présence deux complexes de lanthanide très différents : le $[Eu(HMTDPA)_3]^{3-}$ et le GdHPDO3A. Leurs sites de fixation étant identiques pour les deux protéines modèles, nous voulions étudier le résultat de la compétition entre les deux complexes pour ces sites : l'affinité du complexe $[Eu(HMTDPA)_3]^{3-}$ pour ces protéines est elle en effet, si élevée que ce complexe peut s'imposer au GdHPDO3A ?

Ou observe-t-on davantage une répartition des occupations de sites dans le cristal ?

Nous avons réalisé une série de huit expériences cristallographiques préliminaires avec nos protéines modèles habituelles pour cible : le lysozyme de blanc d'œuf de poule et la thaumatine de *Thaumatococcus daniellii*.

4.7.1 Préparations des différentes combinatoires en fonction du but recherché

Les huit expériences réalisées ont consisté à obtenir des cristaux dérivés de HEWL et TdTHAU avec les huit solutions différentes de complexes de lanthanide hybrides et ce, uniquement par co-cristallisation. Le HEWL a été co-cristallisé avec les huit combinatoires alors que la TdTHAU, seulement avec les combinatoires 5, 6, 7, 8.

La préparation des différentes solutions de combinatoires conduit à un mélange qui contient une distribution de probabilités des différentes combinaisons de ligands qui dépend de ce qui est introduit dans le mélange. Si la constante de dissociation d'un ligand pour l'ion lanthanide est la même pour tous les ligands et si l'échange de ligand a en effet bien lieu en solution, il est alors possible de déplacer cette distribution de probabilités afin de favoriser une des combinaisons. Ceci en modifiant deux paramètres : le nombre de type de ligands introduits dans le mélange et leurs concentrations respectives.

En outre, que les échanges de ligands pour un ion lanthanide soient équiprobables ou non, ces expériences permettent d'émettre des hypothèses à propos de l'affinité des complexes de la combinatoire testée pour la protéine : compte tenu de la compétition pour les sites de fixation, la protéine a-t-elle une affinité particulière pour un complexe ou une combinaison de ligands donnée ?

4.7.2 Présentation des différentes combinatoires et résultats obtenus

Qu'elle que soit la combinatoire, si un doute subsistait quant à la modélisation des ligands dans les cartes de densité électronique, plusieurs affinements ont été menés de conser*ve* avec des modèles contenant les différentes combinaisons de complexes possibles. Dans les tableaux qui vont suivre, les cellules striées de noir signifient que les données de diffraction correspondantes n'ont pas été enregistrées ou se sont révélées non-interprétables.

4.7.2.1 Les combinatoires avec mélange de complexes $[Ln(TDPA)_3]^x$ entiers

Il s'agit de quatre combinatoires réalisées par mélange de solutions contenant chacune un complexe entier (lanthanide + ligands) appartenant à la famille des $[Ln(TDPA)_3]^x$.

Combinatoire 1

Mélange d'une solution de complexe [Eu(HMTDPA)$_3$]$^{3-}$ à 4 mM avec une solution de [Eu(HETDPA)$_3$]$^{3-}$ à 4 mM et d'une solution de Eu(HECTDPA)$_3$ à 4 mM. La solution résultante de ce mélange contient donc, au minimum, 4 mM d'une des combinatoires les moins probables parmi les 10 possibles **[Tableau** 4.23**]**. La solution réalisée a été introduite dans la goutte de cristallisation comme s'il s'agissait d'une solution de complexe unique.

Appellation	
Ligand HMTDPA =	"M"
Ligand HETDPA =	"E"
Ligand HECTDPA =	"C"

Combinaisons		
MMM	EEE	CCC
MME	EEM	CCE
MMC	EEC	CCM
MEE	EMM	CEE
MCC	ECC	CMM
MEC	EMC	CME

Probabilités des différentes combinaisons pour un mélange équimolaire de 3 complexes	
MMM, EEE, CCC	$\frac{1}{18}$
MME, MMC, MEE, MCC, EEC, ECC	$\frac{2}{18}$
MEC	$\frac{3}{18}$

Tableau 4.23 – Probabilité de chacune des différentes combinaisons de ligands rendues possibles par la préparation de la combinatoire 01

Des cristaux dérivés aux faces striées sont apparus dans la goutte de cristallisation composée de 1,33 mM_{t_0} de combinatoire et de 0,46 mM de HEWL. Cassant au niveau des stries, nous avons pu isoler des fragments cristallins qui ont conduit à l'enregistrement de données de diffraction $SAD_{Se_{L_{III}}}$ à une résolution de 1,40 Å. Alors que la solution de combinatoire contenait les ligands HMTDPA, HETDPA et HECTDPA après un phasage expérimental aisé, la carte de densité électronique affinée m'a conduit à modéliser un complexe formé de trois ligands HMTDPA **[Tableau 4.24]**. Un seul des deux sites de fixation du HEWL est occupé à 87 % par le complexe. A l'issue de cette expérience, nous avons pu formuler deux hypothèses :

- ▷ si il n'y a pas échange de ligands, le complexe modélisé est le complexe qui a la plus forte affinité pour la protéine,
- ▷ si il y a échange de ligands, soit il y a sélection des complexes hybrides ayant la plus forte affinité, soit il y a échange de ligands à la surface de la protéine.

Combinatoire 2

Mélange d'une solution de complexe $[Eu(HMTDPA)_3]^{3-}$ à 4 mM avec une solution de $Eu(HECTDPA)_3$ à 8 mM. Suite à ce qui a été observé avec la première combinatoire, celle-ci a été réalisée afin de forcer l'échange de ligands autour de l'ion lanthanide pour favoriser le complexe $Eu(HECTDPA)_3$.

1,33 mM_{t_0} de combinatoire, où la stœchiométrie favorise la formation de complexe avec des ligands HECTDPA, ont été mélangés à 0,46 mM de HEWL. Des cristaux semblables au cristaux de lysozyme natifs sont apparus dans la goutte. Compte tenu de ce que nous avions observé à propos des cristaux indépendamment dérivés avec les complexes $[Eu(HMTDPA)_3]^{3-}$ et $Eu(HECTDPA)_3$, nous nous attendions à ce que ces cristaux contiennent des complexes formés de ligands HECTDPA moins "invasifs" que ceux formés de ligands HMTDPA. Le jeu de données de diffraction MAD à 1,40 Å de résolution m'a cette fois conduit à construire un complexe formé de deux ligands HECTPA et d'un ligand HMTDPA. Cependant, c'est en diminuant le niveau de contour de la carte de densité électronique (à une valeur de 0,70 σ) que la "queue choline" des ligands 1 et 2 apparait de façon distincte. Si les contours sont de 1,0 σ, il m'a été difficile de distinguer la fonction hydroxyethylcholine de la fonction hydroxymethyl des ligands. Modéliser le premier ligand en tant que HECTDPA fut aussi conforté par le fait que la carte de densité électronique affinée ne montre qu'un seul site de fixation occupé à 64 % et une liaison hydrophobe avec le Trp123 de type π-π (caractéristiques du ligand HECTDPA) et non pas deux sites de fixation et une interaction CH-π (caractéristiques du ligand HMTDPA) **[Tableau 4.24]**.

Dans les structures avec un complexe unique, les ligands avaient pu être aisément modélisés. Même observation dans le cas de la combinatoire 4 où le complexe hybride issu d'une synthèse en laboratoire de chimie est bien composé de différents ligands que nous avons pu facilement modéliser. Dans cette mesure, nous pensons qu'il n'y a pas échange de ligands, lorsque l'on mélange des solutions de complexes.

Combinatoire 3

Mélange d'une solution de complexe $[Eu(HMTDPA)_3]^{3-}$ à 4 mM avec une solution de $[Eu(HETDPA)_3]^{3-}$ à 4 mM. En mélangeant seulement deux complexes différents dans des proportions equimolaires alors que ceux-ci sont formés de trois ligands, notre idée était de déterminer quelle est la forme combinée la plus favorable pour une interaction avec la protéine.

Bien que seuls les complexes $[Eu(HMTDPA)_3]^{3-}$ et $[Eu(HETDPA)_3]^{3-}$ qui, normalement, conduisent à la résolution de la structure des deux protéines modèles, aucun cristal au profil de diffraction enregistrable n'a peu être obtenu **[Tableau 4.26]**.

Combinatoire 4

Solution d'un complexe d'europium préparé au laboratoire SEESIB avec un rapport de 1 ligand HMTDPA pour 2 ligands HECTDPA concentrée à 9 mM. Ayant toujours effectué la complexation du lanthanide avec les ligands et réalisé les différents mélanges avec mes propres protocoles, j'ai voulu comparer les résultats obtenus avec lesdits protocoles avec ceux obtenus à partir d'une préparation d'un complexe hybride issue du laboratoire de chimie SEESIB.

3 mM_{t_0} de combinatoire ont été ajoutés au 0,46 mM_{t_0} de HEWL dans la goutte de co-cristallisation. Je n'ai jamais observé de "mauvais cristaux" avec cette combinatoire, seulement des cristaux volumineux aux faces régulières. L'enregistrement de données de diffraction sur ces cristaux a permis d'aboutir à

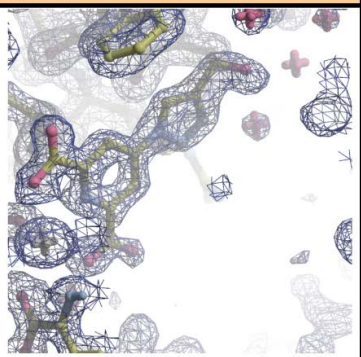
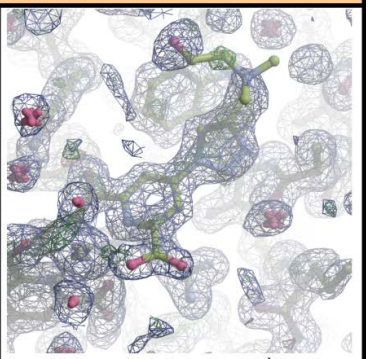
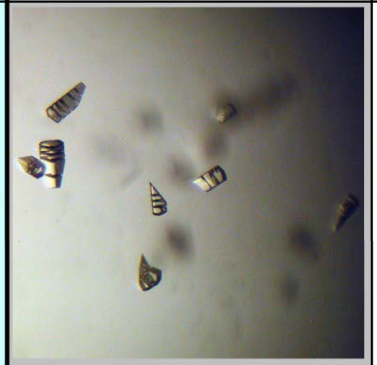
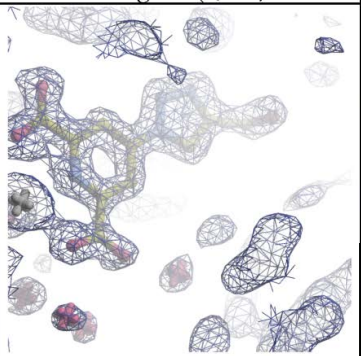
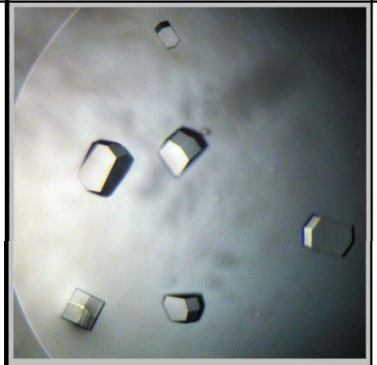
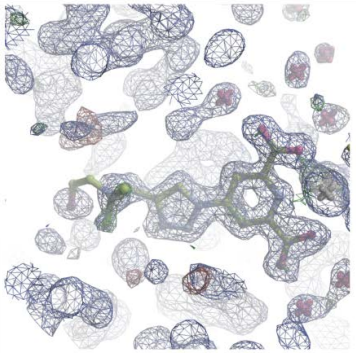
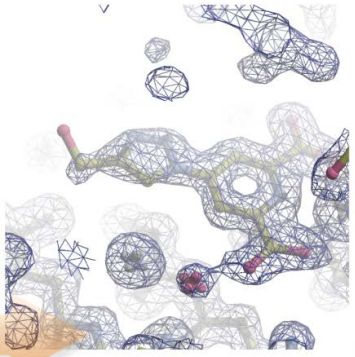
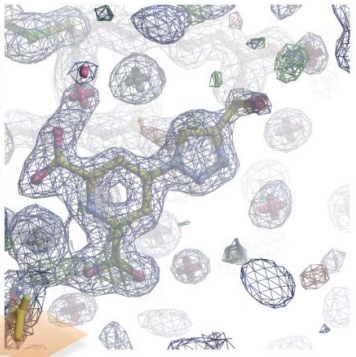
	Combinatoire 1		Combinatoire 2	
Complexe combiné créé / attendu	**1 $[Eu(HMTDPA)_3]^{3-}$ + 1 $[Eu(HETDPA)_3]^{3-}$ + 1 $Eu(HECTDPA)_3$**		**1 $[Eu(HMTDPA)_3]^{3-}$ + 2 $Eu(HECTDPA)_3$**	
But recherché	Combinatoire possible ? Echange de ligands ?		Déplacer la stoechiométrie ? Echange de ligands ?	
Protéine	HEWL		HEWL	
[Complexe] / [Protéine]	2,89		2,89	
Methode d'obtention	co-cristallisation	1er ligand (1,0 σ)	co-cristallisation	1er ligand (0,70 σ)
Faciès cristallin	 *(50 x 80 µm)*	 2d ligand (1,0 σ)	 *(360 x 300 µm)*	 2d ligand (0,70 σ)
$Pics_{Fa}$ (Pic_{Soufre}) pk (σ)			49,47 (4,60)	
$Pics_{Fa}$ (Pic_{Soufre}) rm (σ)	98,79 (4,77)		46,89 (4,46)	
R_{work} (%)	14,98		16,44	
R_{free} (%)	17,06		18;87	
Nbre de réflexions	43878		43532	
Complexe combiné observé	**1 $[Eu(HMTDPA)_3]^{3-}$**	3ème ligand (1,0 σ)	**Mélange $[Eu(HMTDPA)_3]^{3-}$ et $Eu(HECTDPA)_3$**	3ème ligand (0,70 σ)

Tableau 4.24 – Récapitulatif des résultats obtenus avec les combinatoires 1 et 2

la modélisation du complexe. Il occupe, à 65 %, un seul site de fixation. Comme pour la combinatoire 2, la densité électronique affinée des ligands 1 et 2 a conduit à la modélisation de ligands HECTDPA, la "queue choline" étant cette fois visible si les contours de la carte étaient à 0,90 σ. Concernant la modélisation du complexe, la densité électronique relative au premier ligand fut plus ambiguë : elle permit de modéliser à la fois une liaison π-π et CH-π avec le Trp123 **[Tableau 4.25]**.

Le faciès des cristaux obtenus, le taux d'occupation de l'unique site occupé et la faible fixation du complexe sur la TdTHAU sont autant de propriétés qui rappellent celles du complexe $Ln(HECTDPA)_3$ présenté précédemment. En effet, les ligands du complexe $Ln(HECTDPA)_3$ avaient pu être modélisés assez facilement à la surface de la protéine HEWL et le taux d'occupation observé était de 62 %. La synthèse en laboratoire a donc permis de créer, cette fois-ci, un complexe réellement hybride $[(HMTDPA).(HECTDPA)_2]^{1-}$dont la modélisation, contre toute attente, prouve qu'il n'y a pas échange de ligands dipicolinates en solution autour de l'ion lanthanide.

Comme nous l'avons explicité auparavant, la protéine TdTHAU ne permet qu'un mode de fixation impliquant des liaisons hydrogènes avec les acides aminés arginine pour les complexes de la famille des $[(TDPA)_3]^x$. Mode de fixation que nous avons démontré comme dépendant de la charge globale du complexe picolinate utilisé. Dans le cas de cette quatrième combinatoire et bien que le complexe hybride ait permis le phasage, son taux d'occupation n'a pas été assez élevé pour modéliser les ligands. Cela rappelle l'absence de fixation du complexe $Ln(HECTDPA)_3$ sur la thaumatine. Avec deux ligands HECTDPA pour un ligand HMTDPA, la charge globale du complexe n'est plus 3^- mais 1^-. De par son unique charge négative, le complexe hybride $[(HMTDPA).(HECTDPA)_2]^{1-}$ se rapproche du $Ln(HECTDPA)_3$ neutre et ne se fixe que très peu à la TdTHAU.

4.7.2.2 La combinatoire 5, avec complexation de l'ion lanthanide avec deux ligands de la famille $[Ln(TDPA)_3]^x$

Mélange d'une solution contenant 3 mM de ligand HMTDPA, d'une solution contenant 3 mM de ligand HETDPA, d'une solution contenant 3 mM de ligand CTDPA avec une solution de $EuCl_3$ à 3 mM. Cette fois, ce sont les ligands et non plus les complexes complets qui ont été mis en présence de l'ion lanthanide. Nous avons eu cette démarche afin d'augmenter la probabilité d'échange de ligand autour du lanthanide mais aussi afin de tenter de modéliser le ligand CTDPA.

En complexant moi-même trois ligands différents avec les ions europium en proportion equimolaire, je m'attendais à pouvoir modéliser un complexe totalement hybride. Mais le mélange de 1 mM_{t_0} de cette combinatoire avec 0,46 mM_{t_0} de HEWL, n'a permis d'obtenir que des cristaux au faciès torturé. La présence du ligand CTDPA semble être rédhibitoire à l'apparition de cristaux dérivés au profil de diffraction enregistrable **[Tableau 4.26]**.

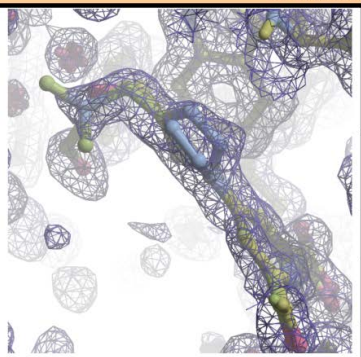
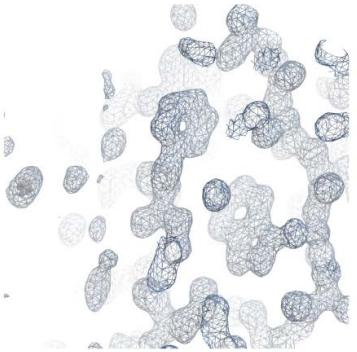
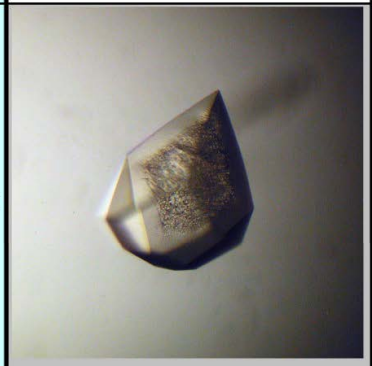
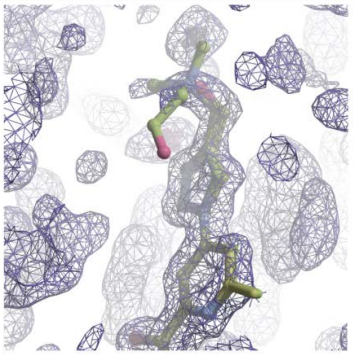
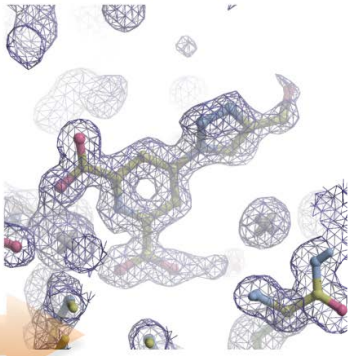
	Combinatoire 4			
Complexe combiné créé / attendu	**Complexe hybride $[Eu(HMTDPA).(HECTDPA)_2]^{1-}$ SEESIB**	 1er ligand (0,90 σ)	**Complexe hybride $[Eu(HMTDPA).(HECTDPA)_2]^{1-}$ SEESIB**	
But recherché	Complexe hybride réalisé en labo de chimie pour comparaison		Complexe hybride réalisé en labo de chimie pour comparaison	
Protéine	HEWL		TdTHAU	
[Complexe] / [Protéine]	6,52		16,67	
Methode d'obtention	co-cristallisation		co-cristallisation	
Faciès cristallin	 *(420 x 550 µm)*	 2d ligand (0,90 σ)	Cristaux obtenus au faciès identique à celui des cristaux de TdTHAU natifs	La protéine TdTHAU ne permet, pour les complexes de la famille des $[(TDPA)_3]^x$, qu'un mode de fixation impliquant des liaisons hydrogène avec les acides aminés arginine. Mode de fixation que nous avons démontré comme dépendant de la charge globale du complexe picolinate utilisé. Bien qu'il ait permis le phasage, le taux d'occupation du complexe hybride n'est pas assez élevé pour modéliser les ligands. De par son unique charge négative, ce complexe hybride se rapproche du $Ln(HECTDPA)_3$ neutre. Il ne peut pas se fixer suffisamment à la TdTHAU.
$Pics_{Fa}$ (Pic_{Soufre}) pk (σ)		 3ème ligand (0,90 σ)	73,13 (18,53)	
$Pics_{Fa}$ (Pic_{Soufre}) rm (σ)	77,37 / 9,17 (4,48)		36,17 (16,57)	
R_{work} (%)	15,56		18,88	
R_{free} (%)	17,21		19,56	
Nbre de réflexions	61403		94347	
Complexe combiné observé	**Mélange $[Eu(HMTDPA)_3]^{3-}$ $Eu(HECTDPA)_3$, $[Eu(HMTDPA)_3]^{3-}$ major.**		**$Eu(HECTDPA)_3$ majoritaire ?**	

Tableau 4.25 – Récapitulatif des résultats obtenus avec la combinatoire 4

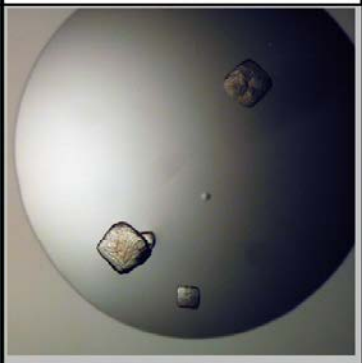
	Combinatoire 3	Combinatoire 5	
Complexe combiné créé / attendu	1 $[Eu(HMTDPA)_3]^{3-}$ + 1 $[Eu(HETDPA)_3]^{3-}$	1 HMTDPA + 1 HETDPA + 1 CTDPA + 3 Eu	
But recherché	Quelle combinatoire majoritaire / favorable ?	Combinatoire facilitée si ligands utilisés ?	
Protéine	HEWL	HEWL	TdTHAU
[Complexe] / [Protéine]	2,89	2,89	2,22
Methode d'obtention	co-cristallisation	co-cristallisation	co-cristallisation
Faciès cristallin	 *(280 × 280 µm)*	*(300 × 220 µm)*	Cristaux obtenus au faciès identique à celui des cristaux de TdTHAU natifs
$Pics_{Fa}$ (Pic_{Soufre}) pk (σ)			Aucune fixation de complexe de lanthanide détectée.
$Pics_{Fa}$ (Pic_{Soufre}) rm (σ)			
R_{work} (%)			
R_{free} (%)			
Nbre de réflexions			
Complexe combiné observé			

Tableau 4.26 – Récapitulatif des résultats obtenus avec les combinatoires 3 et 5

4.7.2.3 Les combinatoires avec mélange de complexes entiers appartenant à la grande famille des complexes picolinates

Il s'agit de deux combinatoires réalisées par mélange de solutions contenant chacune le complexe $[Eu(HMTDPA)_3]^{3-}$ et le complexe $[Eu(DPA)_3]^{3-}$.

Combinatoire 6

Mélange d'une solution de complexe $[Eu(HMTDPA)_3]^{3-}$ à 15 mM avec une solution de $[Eu(DPA)_3]^{3-}$ à 15 mM. Tentative pour créer un complexe hybride ayant un ou plusieurs ligands dipicolinates avec un ou plusieurs ligands HMTDPA. Le but étant de voir quel faciès ont les cristaux dérivés obtenus par co-cristallisation et d'identifier de nouvelles interactions supramoléculaires.

5 mM_{t_0} d'une solution contenant le mélange de complexes $[Eu(HMTDPA)_3]^{3-}$ et $[Eu(DPA)_3]^{3-}$ ont été mis en présence de 0,46 mM de HEWL. Des cristaux quatratiques aux faciès torturés sont apparus,

semblables à ceux que l'on obtient lorsqu'on co-cristallise la protéine avec une concentration en complexe $[Eu(DPA)_3]^{3-}$ inférieure à 50 mM **[Tableau 4.27]**. Aucune donnée de diffraction interprétable n'a pu être enregistrée. On devine bien ici l'influence du complexe $[Eu(DPA)_3]^{3-}$ qui ne peut que provoquer un effet supramoléculaire partiel conduisant aux "mauvais cristaux".

Combinatoire 7

Mélange d'une solution de complexe [Eu(HMTDPA)$_3$]$^{3-}$ à 15 mM avec une solution de [Eu(DPA)$_3$]$^{3-}$ à 150 mM. Tentative pour créer un complexe hybride ayant un ou plusieurs ligands dipicolinates avec un ou plusieurs ligands HMTDPA. Car c'est à partir de 50 mM que le complexe $[Eu(DPA)_3]^{3-}$ conduit usuellement à la forme monoclinique des cristaux dérivés de HEWL, la stœchiométrie est ici en faveur du complexe $[Eu(DPA)_3]^{3-}$, ceci afin de déterminer quel complexe a le plus d'affinité pour cette protéine.

Par rapport à la sixième combinatoire, le complexe $[Eu(DPA)_3]^{3-}$ a été ajouté dans le mélange à une concentration suffisante pour conduire à la forme monoclinique des cristaux de lysozyme. Dans le cadre de cette expérience, le complexe $[Eu(HMTDPA)_3]^{3-}$est donc largement minoritaire. Ainsi, bien que les cristaux dérivés obtenus aient des faces plus régulières que ceux obtenus avec la combinatoire 6, aucun jeu de données de diffraction interprétable n'a pu être enregistré **[Tableau 4.27]**. Ceci prouve que même en faible quantité, le complexe $[Eu(HMTDPA)_3]^{3-}$ a une telle affinité pour la protéine qu'il peut entrer en compétition avec le $[Eu(DPA)_3]^{3-}$au point d'empêcher l'effet supramoléculaire $[Eu(DPA)_3]^{3-}$(ce dernier étant pourtant introduit en large excès).

4.7.2.4 La combinatoire 8, avec mélange de deux complexes dissemblables : le $[Eu(HMTDPA)_3]^{3-}$ et le GdHPDO3A

Mélange d'une solution de complexe [Eu(HMTDPA)$_3$]$^{3-}$ à 15 mM avec une solution de complexe GdHPDO3A à 240 mM. Combinatoire qui met en compétition les deux complexes ayant une forte affinité pour le HEWL et la TdTHAU.

En utilisant conjointement les complexes $[Eu(HMTDPA)_3]^{3-}$ et GdHPDO3A à des concentrations suffisantes pour qu'il se fixent à la protéine (respectivement 5 mM_{t_0} et 80 mM_{t_0}), nous provoquons donc une compétition de site entre ces deux complexes radicalement différents. Lors de l'enregistrement de données de diffraction des cristaux dérivés de HEWL avec ce mélange, se placer au seuil d'absorption L_{III} de l'europium (1,774 Å) éteint partiellement le signal anomal du gadolinium (car son seuil d'absorption L_{III} est, lui, de 1,709 Å). Ainsi, au seuil L_{III} du Gd, le signal anomal du Eu et du Gd sont détecté (f'' de 10 e^- et 30 e^-, respectivement) alors qu'au seuil L_{III} du Eu, seul le signal anomal du Eu reste important (f'' de 30 e^- pour le Eu et 4 e^- pour le Gd). On voit ici trois pics intenses en Fourier anomale au seuil du gadolinium, prouvant la fixation majoritaire du complexe GdHPDO3A **[Tableau 4.28]**.

Encore une fois, bien que des cristaux de thaumatine co-cristallisés avec cette combinatoire aient été obtenus, aucun des traitements de données réalisés n'a permis de détecter la fixation d'un des diffuseurs anomaux.

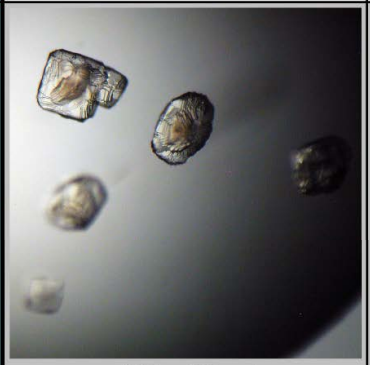
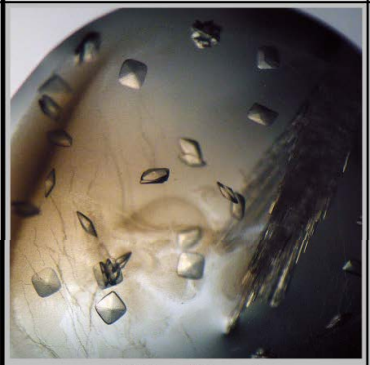
	Combinatoire 6		Combinatoire 7	
Complexe combiné créé / attendu	**1 $[Eu(HMTDPA)_3]^{3-}$ + 1 $[Eu(DPA)_3]^{3-}$**		**1 $[Eu(HMTDPA)_3]^{3-}$ + 10 $[Eu(DPA)_3]^{3-}$**	
But recherché	Compatible avec supra ?		Compétition d'affinité	
Protéine	HEWL	TdTHAU	HEWL	TdTHAU
[Complexe] / [Protéine]	10,87	2,78	Dépend de la compétition	Dépend de la compétition
Methode d'obtention	co-cristallisation	co-cristallisation	co-cristallisation	co-cristallisation
Faciès cristallin	 *(300 x 220 µm)*	Cristaux obtenus au faciès identique à celui des cristaux de TdTHAU natifs	 *(180 x 180 µm)*	Cristaux obtenus au faciès identique à celui des cristaux de TdTHAU natifs
$Pics_{Fa}$ (Pic_{Soufre}) pk (σ)		Aucune fixation de complexe de lanthanide détéctée.		Aucune fixation de complexe de lanthanide détéctée.
$Pics_{Fa}$ (Pic_{Soufre}) rm (σ)				
R_{work} (%)				
R_{free} (%)				
Nbre de réflexions				
Complexe combiné observé				

Tableau 4.27 – Récapitulatif des résultats obtenus avec les combinatoires 6 et 7

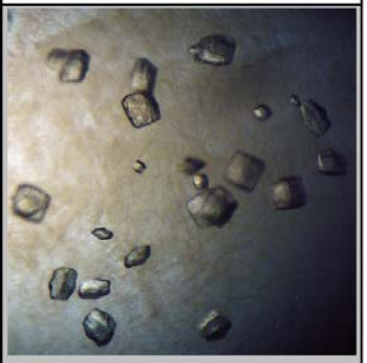
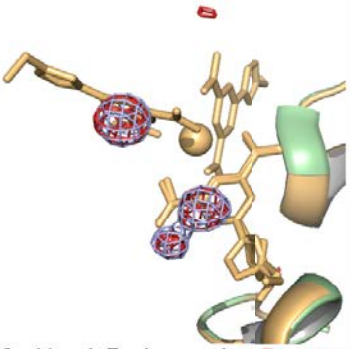
	Combinatoire 8		
Complexe combiné créé / attendu	1 $[Eu(HMTDPA)_3]^{3-}$ + 16 GdHPDO3A	Se placer au seuil d'absorption L_{III} du Eu (1,774 Å) éteint partiellement le signal anomal du Gd (seuil d'absorption L_{III} du Gd : 1,709 Å). Ainsi, au seuil d'absorption L_{III} du Gd, le signal anomal du Eu et du Gd est détecté alors qu'au seuil d'absorption L_{III} du Eu, seul le signal anomal du Eu est détecté. $f''_{Eu_{L_{III}}Gd} = 10\ e^-$ et $f''_{Gd_{L_{III}}Eu} = 4\ e^-$	1 $[Eu(HMTDPA)_3]^{3-}$ + 16 GdHPDO3A
But recherché	Compétition de site		Compétition de site
Protéine	HEWL		TdTHAU
[Complexe] / [Protéine]	Dépend de la compétition		Dépend de la compétition
Methode d'obtention	co-cristallisation		co-cristallisation
Faciès cristallin	 *(160 × 160, 40 × 40 µm)*	 Synthèses de Fourier anomale au $Eu_{L_{III}}$ et au $Gd_{L_{III}}$ (cartes contourées à 5 σ). Modèles superposés de HEWL+Eu[$(HMTDPA)_3]^{3-}$ et de HEWL+GdHPDO3A	Cristaux obtenus au faciès identique à celui des cristaux de TdTHAU natifs
$Pics_{Fa}$ (Pic_{Soufre}) Eu pk (σ)	27,97 / 23,70 (6,20)		
$Pics_{Fa}$ (Pic_{Soufre}) Gd pk (σ)	44,50 / 35,67 / 16,15 (4,01)		
R_{work} (%)			
R_{free} (%)			
Nbre de réflexions			Aucune fixation de complexe de lanthanide détéctée.
Complexe combiné observé	Davantage de GdHPDO3A que de $[Eu(HMTDPA)_3]^{3-}$		

Tableau 4.28 – Récapitulatif des résultats obtenus avec la combinatoire 8

Cette fois-ci l'affinité du complexe $[Eu(HMTDPA)_3]^{3-}$ pour le lysozyme n'a pu rivaliser avec celle du GdHPDO3A ajouté en large excès. C'est ce que montrent les synthèses de Fourier anomale avec les deux modèles, HEWL+$[Eu(HMTDPA)_3]^{3-}$ et HEWL+GdHPDO3A (Girard *et al.*, 2003b), alignés et superposés du **tableau 4.28** et de la **figure 4.50.** On voit les deux pics relatifs au gadolinium à l'emplacement exact du modèle superposé qui prouve que le complexe GdHPDO3A est bien le complexe fixé. Les pics symétriques visibles au niveau du modèle du complexe $[Eu(HMTDPA)_3]^{3-}$ montrent bien que ce dernier ne s'est que peu ou pas fixé. La très faible fixation du $[Eu(HMTDPA)_3]^{3-}$ pourrait cependant expliquer le troisième pic observable par la synthèse de Fourier anomale calculée avec le jeu gadolinium. Il pourrait s'agir de la diffusion anomale des atomes de gadolinium de complexes GdHPDO3A qui se sont décalés a cause de la présence du $[Eu(HMTDPA)_3]^{3-}$ présents dans la même région.

Dans le cas des combinatoires 5, 6, 7 et 8, l'absence de fixation dans les cristaux de TdTHAU peut s'expliquer par le fait que nous avons utilisé des solutions de combinatoires à des concentrations trop faibles pour obtenir des cristaux dérivés de cette protéine.

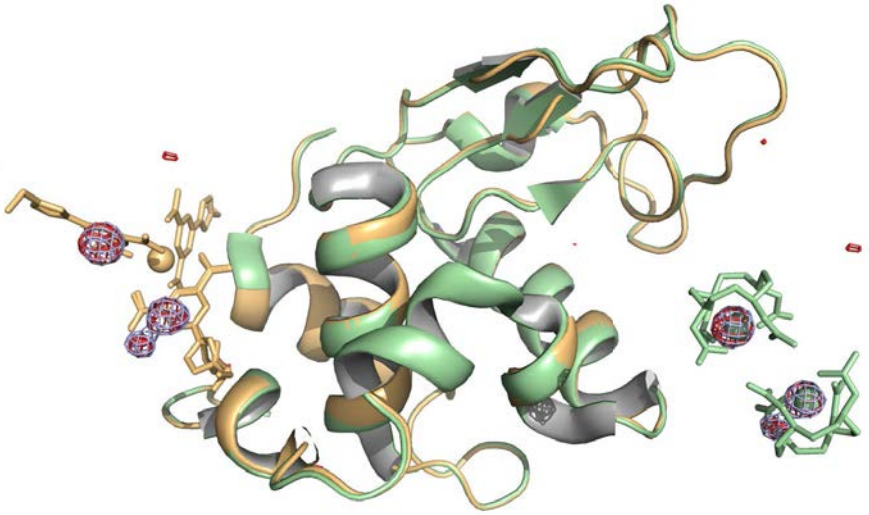


Figure 4.50 – Synthèses de Fourier anomale calculées avec le jeu de données SAD enregistré au seuil d'absorption L_{III} du gadolinium (en bleu-gris) et avec le jeu de données SAD enregistré au seuil d'absorption L_{III} de l'europium (en rouge). Les modèles de HEWL avec le complexe $[Eu(HMTDPA)_3]^{3-}$ (en orangé) et de HEWL avec le GdHPDO3A (en vert) sont alignés et superposés

4.8 Conclusions quant à l'utilisation des complexes de la famille des $[Ln(TDPA)_3]^x$

L'utilisation du complexe $[Ln(HMTDPA)_3]^{3-}$, dérivé chimique du $[Ln(DPA)_3]^{3-}$, a conduit à l'obtention de phases très précises qui ont permis la résolution de la structure du lysozyme de blanc d'œuf de poule et de la thaumatine de *Thaumatococcus daniellii*, deux protéines modèles.

Son affinité pour la surface de ces protéines est telle qu'elle permet de les utiliser à une concentration de quelques millimolaires seulement. La forte affinité de ce complexe et des autres membres de la famille des $[Ln(TDPA)_3]^x$ est liée à la présence du noyau triazole qui crée des interactions de type CH-π et empliement-π avec le noyau indole des acides aminés tryptophane de la surface protéique. Ce groupement, qui rend donc les complexes $[Ln(TDPA)_3]^x$ plus affins que le complexe $[Ln(DPA)_3]^{3-}$, ne les empêche pas d'établir aussi des liaisons hydrogènes avec les acides aminés arginine et lysine, mode de fixation préféré du complexe $[Ln(DPA)_3]^{3-}$.

Leur luminescence dans le visible, lorsqu'excité sous UV proches, permet de le détecter aisément dans les cristaux dérivés de protéines et ce, même si le complexe est utilisé à basse concentration.

Chacun des complexes qui compose la famille des $[Ln(TDPA)_3]^x$ est un dérivé du $[Ln(HMTDPA)_3]^{3-}$ et ne diffère que par le groupement chimique ajouté au noyau triazole. La possibilité de greffer par chimie-clic un noyau triazole portant une fonction particulière a permis de créer différents ligands picolinates. Cette panoplie de ligands permet à son tour de composer différents complexes de lanthanide formés soit de trois ligands identiques, soit de ligands différents. Cela confère à chacun de ces complexes uniques des propriétés cristallographiques différentes qui offrent donc la possibilité de choisir le complexe selon la composition en acides aminés de la protéine ciblée.

L'étude des complexes uniques et des complexes hybrides de la famille des $[Ln(TDPA)_3]^x$ a aussi révélé, contre toute attente, qu'il n'y a pas échange de ligands dipicolinates en solution autour de l'ion lanthanide.

Chapitre 5

Un tris-dipicolinate cagé, le LnTNTPA

Les membres de la famille des complexes tris-triazoledipicolinates de lanthanide ont permis d'obtenir des phases très précises pour résoudre la structure à haute résolution de deux protéines modèles et ce, même à très basse concentration : le lysozyme de blanc d'œuf de poule et la thaumatine de *Thaumatococcus daniellii*. Ils ont, en outre, ouvert la voie à des complexes modulables. Mais bien que plus luminescents encore que le tris-dipicolinate de lanthanide dont ils sont issus, ils conservent cependant son défaut : ils ne peuvent être utilisés en présence d'ions divalents. C'est dans ce contexte qu'a été synthétisé un nouveau complexe, toujours dérivé du tris-dipicolinate de lanthanide : le LnTNTPA.

5.1 Description du complexe de lanthanide LnTNTPA

Dans ce complexe, les trois ligands dipicolinates sont reliés par un cycle à trois azotes et six carbones (tri-azacyclononane) formant ainsi une cage chimique qui permet d'avoir les neuf coordinations nécessaires à la liaison d'un atome de lanthanide **[Figure 5.1]**. Sa synthèse revient à Olivier Maury, collaborateur du Laboratoire de Chimie de l'Ecole Normale Supérieure de Lyon. Contrairement au tris-dipicolinate de lanthanide ou au tris-hydroxyméthyltriazoledipicolinate, la charge globale de ce complexe est nulle. Le cycle tri-azacyclononane et cette charge globale nulle sont des caractéristiques qui rappellent les complexes destinés à l'imagerie médicale (LnHPDO3A, etc.). Cependant, grâce aux trois acides picoliniques qui viennent compléter la sphère de coordination de l'atome de lanthanide, ce nouveau complexe conserve toutes les propriétés de luminescence. Aucune molécule d'eau ne vient alors masquer l'émission de fluorescence.

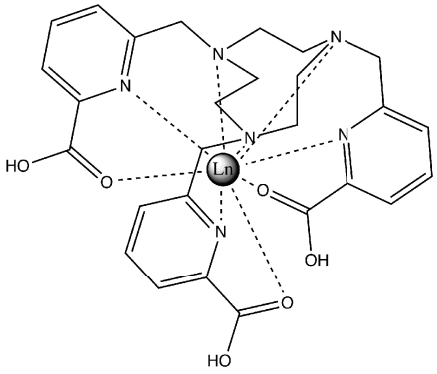


Figure 5.1 – Formule développée du 6,6',6"-(1,4,7-triazonane-1,4,7-triyl)tris(methylène)tripicolinate de lanthanide : "LnTNTPA"

Le complexe m'a été fourni sous la forme de poudre, complexé avec le terbium. La masse molaire de la cage seule est de 666,46 g.mol^{-1}. Celle du complexe de terbium est de 825,38 g.mol^{-1}. Le complexe est très soluble dans l'eau (74 mM à 25°C). Tout comme les complexes précédents, le même protocole de préparation est applicable et on peut vérifier la solubilisation du complexe en exposant la solution aux UV proches (λ = 315 nm).

La synthèse du ligand TNTPA revient à greffer les trois acides dipicoliniques sur un cycle triazacyclononane par alkylation et via l'utilisation d'une de leur fonction acide. L'adjonction du lanthanide se fait ensuite par adjonction du chlorure correspondant et en présence de carbonate de potassium **[Figure 5.2]**.

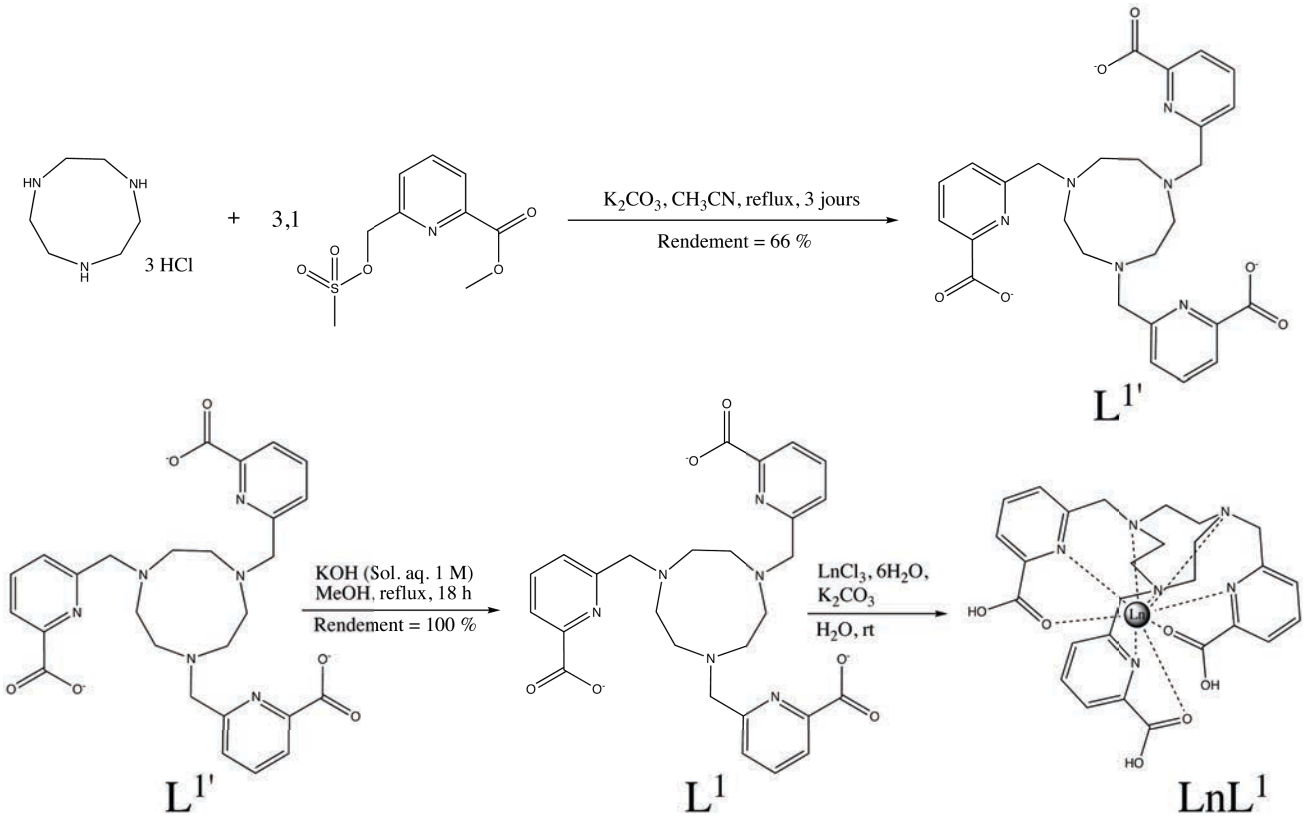


Figure 5.2 – Synthèse chimique du TNTPA par alkylation et complexation du lanthanide

A l'issue de cette synthèse chimique, les résultats spectroscopiques confirment la complexation du lanthanide et permettent de mesurer la luminescence dans le visible. La luminescence du LnTNTPA mesurée avec l'europium est supérieure à celles des complexes tris-dipicolinates d'europium et tris-hydroxymethyltriazoledipicolinate d'europium (τ = 0,83 ms), le temps de demi-vie mesuré (τ) étant de 1,09 ms avec un ion Eu^{3+} **[Figure 5.3 page suivante]**.

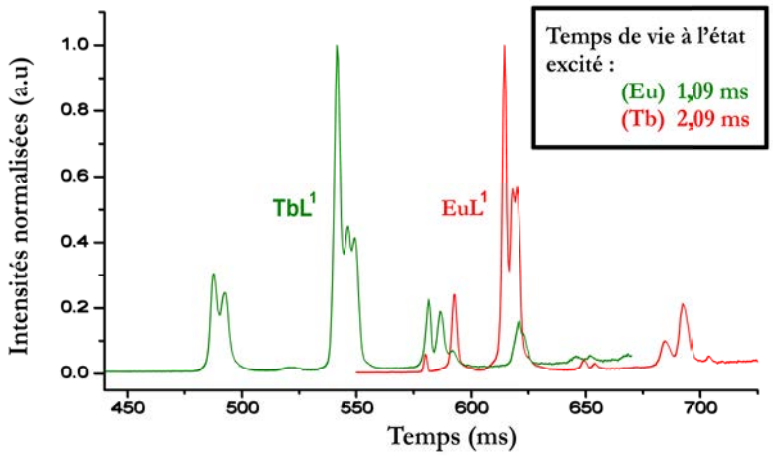
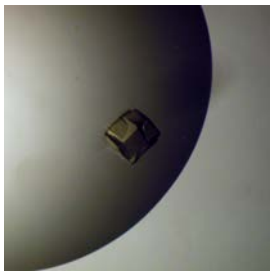


Figure 5.3 – Résultats spectroscopiques à l'issue de la synthèse du TbTNTPA et du EuTNTPA

J'ai testé ce nouveau complexe en co-cristallisation et en trempage avec nos deux protéines modèles habituelles : le lysozyme de blanc d'œuf de poule et la Thaumatine de *Taumatococcus daniellii*. Mais comme la cage qui forme le complexe a pour but d'empêcher tout ion divalent de déstabiliser le complexe, nous nous devions de trouver une protéine modèle cristallisant en présence de ce type d'ions. Le complexe étant de surcroit de taille plus importante, notre choix s'est porté sur la malate déshydrogénase de *Chloroflexus aurantiacus* dont les cristaux apparaissent en présence de 40 mM d'acétate de cadmium et dont les diamètres des canaux de solvant sont importants (*cf.* chapitre 2 de la seconde partie).

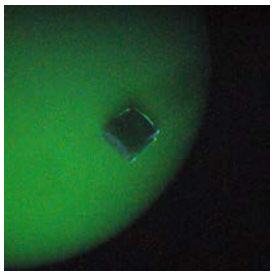
5.2 Co-cristallisation du HEWL avec le complexe TbTNTPA : pas de fixation du TbTNTPA

Bien que des cristaux de lysozyme de blanc d'œuf de poule soient apparus en co-cristallisation avec 24 mM_{t_0}, 30 mM_{t_0} et 60 mM_{t_0}, et malgré l'enregistrement de données MAD à trois longueurs d'onde, je n'ai pas pu constaté de fixation du complexe TbTNTPA. Aucun signal anomal n'est présent dans les données de diffraction. La synthèse de Fourier anomale ne permet pas non plus de détecter des sites de fixation. L'absence de luminescence des cristaux, de formes identiques à celles des cristaux natifs, par rapport au reste de la goutte corrobore cela **[Figure 5.4]**.

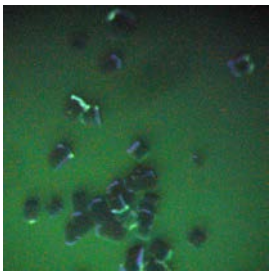


(a) HEWL + 24 mM TbTNTPA, lumière visible

(b) HEWL + 24 mM TbTNTPA, lumière visible



(c) HEWL + 24 mM TbTNTPA, UV proches



(d) HEWL + 24 mM TbTNTPA, UV proches

Figure 5.4 – Cristaux non-luminescents de HEWL obtenus en présence de 24 mM de TbTNTPA

Le complexe LnTNTPA ne semble donc pas capable d'effet supramoléculaire. La compacité de l'empilement cristallin dans les cristaux de lysozyme de groupe d'espace $P4_32_12$ (27 % de solvant) nous montre peut-être ici que nous avons atteint une taille de complexe qui, dans ce cas, ne permet pas sa fixation.

5.3 Co-cristallisation de la TdTHAU avec le TbTNTPA : fixation du TbTNTPA

Contrairement au lysozyme, les cristaux de TdTHAU dérivés par co-cristallisation avec le TbTNTPA, ce dernier ajouté à une concentration de 31 mM_{t_0}, ont conduit à la résolution de la structure de la protéine par phasage expérimental. Mais malgré la haute résolution, le ligand n'a pu être modélisé car le taux d'occupation du complexe est modérément élevé.

5.3.1 Obtention de cristaux dérivés de TdTHAU avec le TbTNTPA

Les cristaux de thaumatine co-cristallisés en présence de plus de 30 mM de complexe TbTNTPA ont conduit à la résolution de la structure de la protéine. Les cristaux dérivés sont semblables aux cristaux natifs de thaumatine. A l'inverse, aucune fixation du complexe dans les cristaux issus de trempages dans des solutions concentrées à 31, 40 ou 70 mM en TbTNTPA n'a été constatée.

5.3.2 Intégration des données de diffraction issues d'un cristal de TdTHAU obtenu par co-cristallisation avec le TbTNTPA

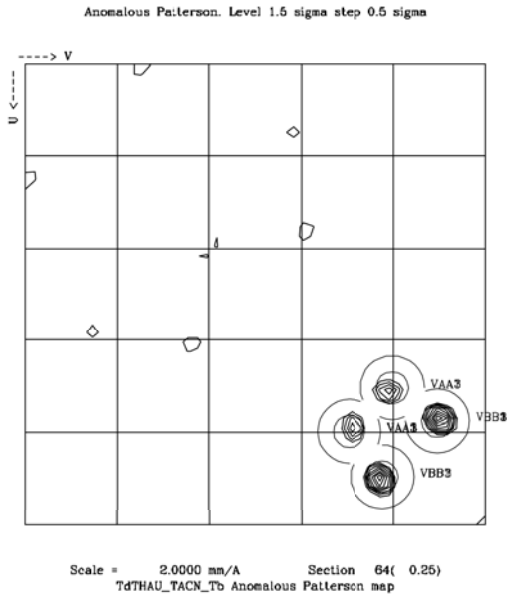
Suite au traitement des données MAD, on détecte un signal anomal un peu moins élevé que ce que nous avions pu constater avec les membres de la famille $[Ln(TDPA)_3]^x$.

	TbTNTPA	
	TdTHAU	
[Complexe]$_{t_0}$ (mM)	31	
[Protéine]$_{t_0}$ (mM)	0,6	
Rapport	51,6	
Méthode	co-cristallisation	
Source synchrotron	ESRF	
Ligne de lumière	FIP-BM30A	
MAD	oui	
λ (Å)	1,648 ($Tb_{L_{III}}$)	0,979
Groupe d'espace	$P4_12_12$	
Paramètres de maille (Å,°)	a = 57,880, c = 150,123	
Contenu unité asymétrique	1 molécule	
Résolution (Å)	45,84 - 1,79 (1,89 - 1,79)	45,89 - 1,30 (1,37 - 1,30)
Nombre de réflexions uniques	24832 (3508)	64086 (9183)
R_{merge} (%)	4,3 (8,2)	5,4 (42,7)
R_{pim} (%)	1,7 (3,5)	2,1 (17,0)
Sig_{ano} XDS	1,55 (1,04)	0,86 (0,80)
R_{ano} (%)	2,3 (3,7)	1,7 (11,8)
$\frac{I}{\sigma(I)}$	11,7 (8,9)	9,5 (1,8)
Complétude (%)	99,9 (99,5)	100,0 (99,9)
Redondance	13,1 (12,4)	13,9 (13,8)

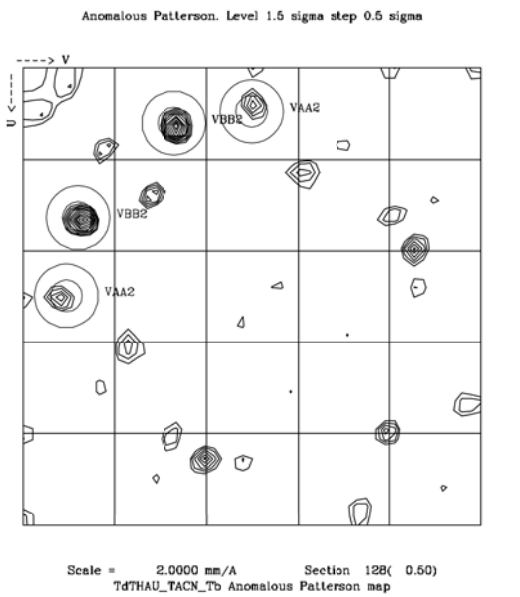
Tableau 5.1 – Statistiques d'intégration du jeu de donnée de diffraction d'un cristal dérivés TdTHAU + TbTNTPA

5.3.3 Phasage expérimental de la TdTHAU avec le TbTNTPA

Les données MAD (enregistrées au seuil L_{III} du terbium et hors seuil) ont permis de mettre en évidence la présence de deux diffuseurs anomaux dans l'unité asymétrique **[Figure 5.5]**. Le jeu de phases expérimentales obtenu permit au programme *ArpWARP* de construire automatiquement un modèle : 171 acides aminés sur 207 ont pu être automatiquement construits à partir de la carte expérimentale améliorée par aplatissement de solvant **[Figure 5.6]**.



(a) Carte de Patterson anomale, section de Harker w = $\frac{1}{4}$



(b) Carte de Patterson anomale, section de Harker w = $\frac{1}{2}$

Figure 5.5 – Deux sections de la carte de Patterson anomale du cristal dérivé de TdTHAU par co-cristallisation avec 31 mM_{t_0} du TbTNTPA (jeu de données MAD)

	TbTNTPA
	TdTHAU
Type de données	MAD
Résolution (Å)	40,83 - 1,30
Pics$_{Fa}$ (Pic$_{Soufre}$) pk (σ)	58,28 / 52,75 (13,81)
Pics$_{Fa}$ (Pic$_{Soufre}$) rm (σ)	25,72 / 21,83 (8,89)
FOM$^{acen.}_{sharp}$ / FOM$^{cen.}_{sharp}$	0,23 / 0,06
Nombre de sites	2
Résolution FOM$^{acen.}_{sharp}$ ≈ 0,5 (Å)	3,35
P$\varphi^{\neq isomorphes}_{sharp\,acen.}$ / P$\varphi^{\neq anomales}_{sharp\,acen.}$	0,00 / 0,40
SIGMAA$_{solomon}$	0,52
Arp/Warp (construit / total)	171 / 207

Tableau 5.2 – Statistiques du phasage expérimental de la TdTHAU avec le TbTNTPA

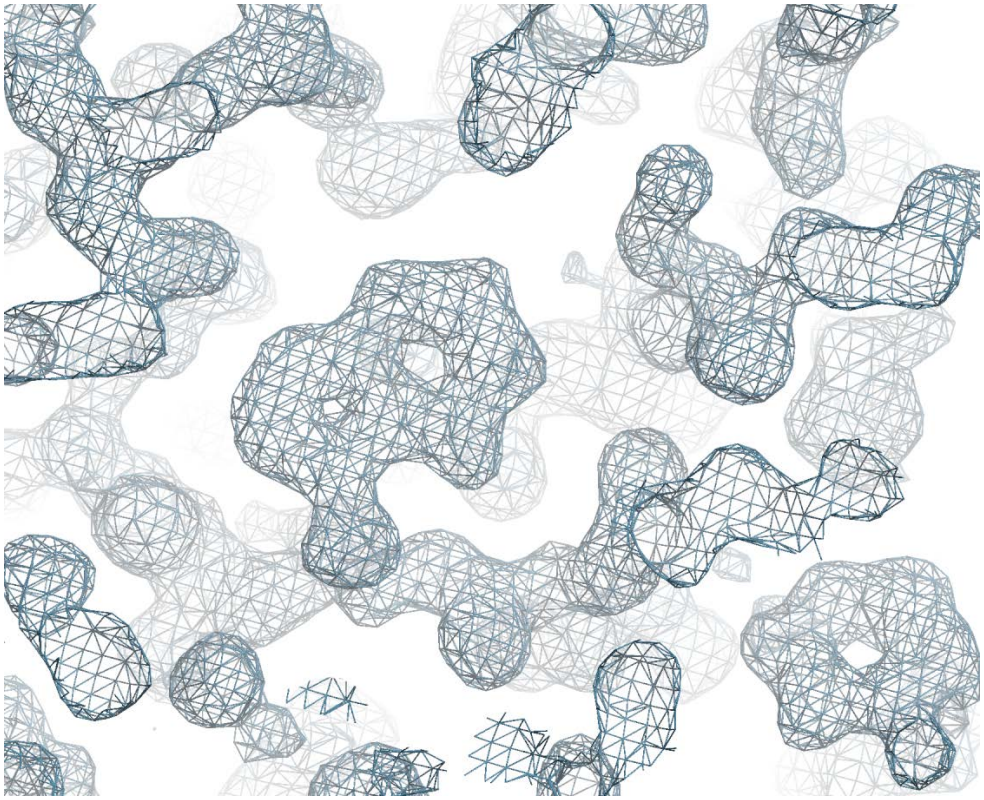


Figure 5.6 – Carte de densité électronique **expérimentale** d'un cristal dérivé de TdtHAU + TbTNTPA (contourée à 1,0 σ) améliorée par aplatissement de solvant. Phasage MAD, résolution : 1,30 Å

5.3.4 Affinement du modèle et description du mode de fixation du TbTNTPA à la surface de la TdTHAU

Deux sites de fixation se trouvent à la surface de la protéine. Mais, hélas, à part celle de l'atome de lanthanide, aucune densité électronique correspondant au ligand-cage n'est visible, que ce soit dans l'un ou l'autre des sites. Ainsi, seuls les atomes de lanthanide ont été placés. Nous ne pourrons donc qu'intuiter un mode de fixation pour ce nouveau complexe. Fidèle à la lignée des dipicolinates, les deux sites se trouvent dans une zone chargée positivement, charge notamment due à la présence de deux acides aminés arginine par site.

	TbTNTPA
	$TdTHAU_{co-cristallisation}$
[Complexe] (mM)	31
Résolution (Å)	40,83 - 1,30
R_{work} (%)	13,81
R_{free} (%)	15,99
Nbre de réflexions utilisées	63889
Composition (nbre d'atomes)	
Protéine	1588
H2O	396
Ions	2
Ligands	20
Res. hors Ramachandran (%)	0
Ecarts types (globaux)	
$Liaisons_{Longueurs}$ (Å)	0,019
$Liaisons_{Angles}$ (°)	0,250
B_{moyen} protéine (Å^2)	12,55
B_{min} protéine (Å^2)	5,49
B_{max} protéine (Å^2)	47,26
B_{moyen} eau (Å^2)	29,23
B_{moyen} ions lanthanide (Å^2)	18,38
B_{moyen} ligands (Å^2)	-
Site 1 $_{Occ}$ (%)	20
Site 2 $_{Occ}$ (%)	15

Tableau 5.3 – Résumé des statistiques d'affinement pour la structure de la TdTHAU avec le TbTNTPA

5.3.4.1 Description du site de fixation 1 (taux d'occupation : 20 %)

Le premier site se situe dans la même cavité que celle décrite dans le chapitre relatif à la fixation des complexes $[Ln(TDPA)_3]^x$ (chapitre précédent). Il s'agit de la même cavité, chargée positivement et formée par l'empilement cristallin dans laquelle vient se fixer la plupart des complexes de lanthanide chargés **[Figure 5.8]**. Ici aussi, encore une fois, la chaine latérale de l'Arg79 vient "refermer" cette cavité, y piégeant le complexe. Cette fois, l'atome de lanthanide ne se situe pas tout à fait sur l'axe de symétrie d'ordre 2, comme cela avait été le cas avec le complexe $[Ln(HETDPA)_3]^{3-}$, ce décalage a d'ailleurs permis le phasage expérimental direct **[Figure 5.7]**. Les taux d'occupation du site de référence et de son symétrique sont donc de 20 %.

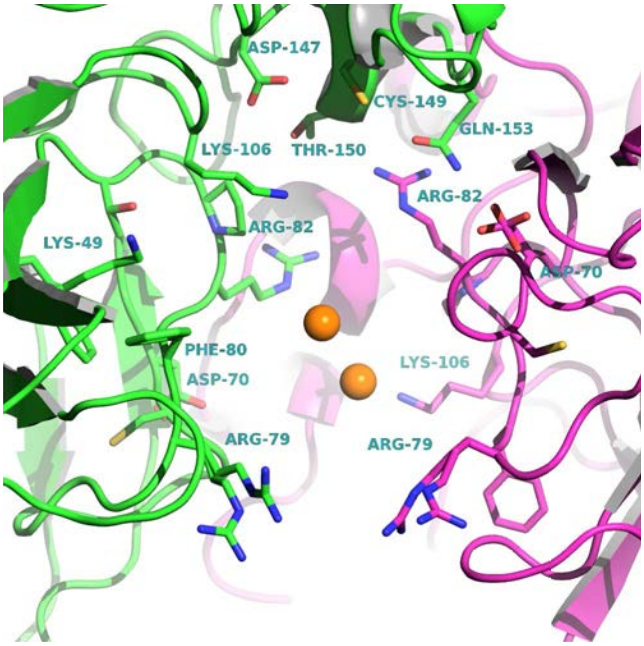


Figure 5.7 – Premier site de fixation du TbTNTPA à une interface entre deux molécules de TdTHAU. Cristal dérivé obtenu par co-cristallisation. La molécule de référence est en vert, la molécule symétrique, en rose

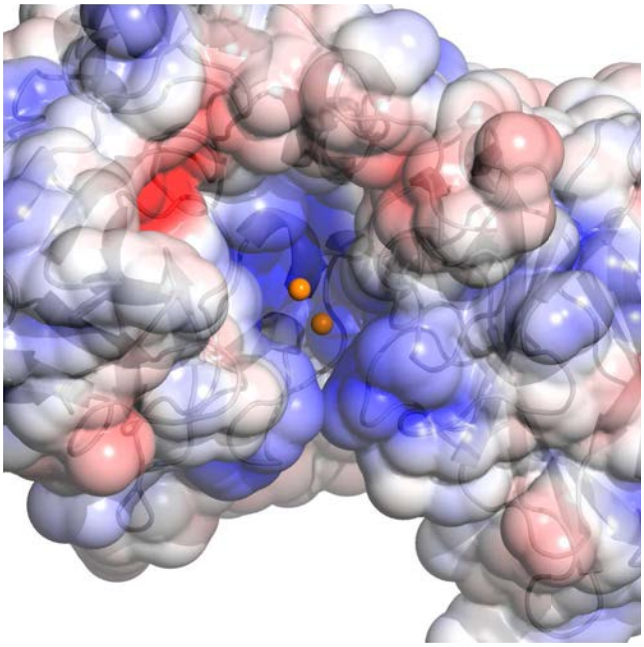


Figure 5.8 – Surface électrostatique du premier site de fixation du TbTNTPA. Echelle allant de - 3,0 kT.e^{-1} (charge négative, rouge) à + 3,0 kT.e^{-1} (charge positive, bleu)

Alors que tous les complexes de la grande famille des complexes picolinates occupent exclusivement ce site de fixation, le complexe LnTNTPA se fixe dans un deuxième site à la surface de la protéine TdTHAU, en dehors de cette cavité.

5.3.4.2 Description du site de fixation 2 (taux d'occupation : 15 %)

Le second site de fixation est, lui aussi, à l'interface de deux molécules dans le cristal **[figure 5.9]**. Il montre que les deux chaines latérales des acides aminés Ag119 et Arg122 de la molécule de référence s'écartent comme pour laisser la place à quelque chose d'imposant. La **figure 5.10** montre que la charge globale de ce site de fixation est bien plus neutre que celle du site majoritairement peuplé.

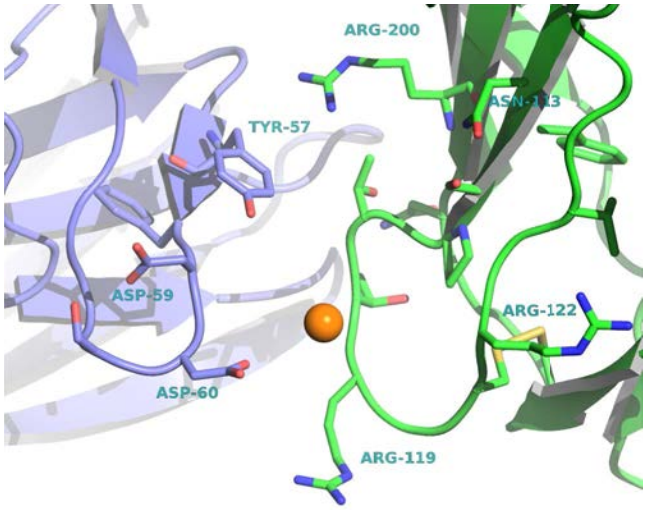


Figure 5.9 – Second site de fixation du TbTNTPA à une interface entre deux molécules de TdTHAU. Cristal dérivé obtenu par co-cristallisation. La molécule de référence est en vert, la molécule symétrique, en bleu

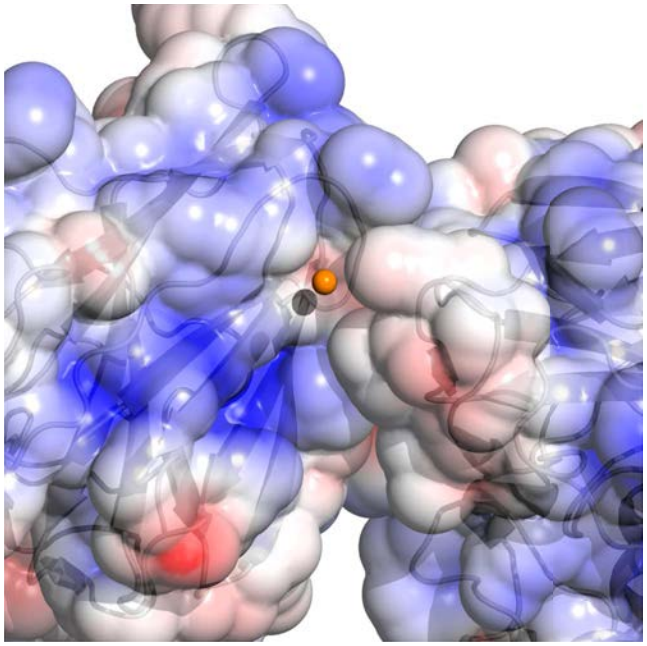
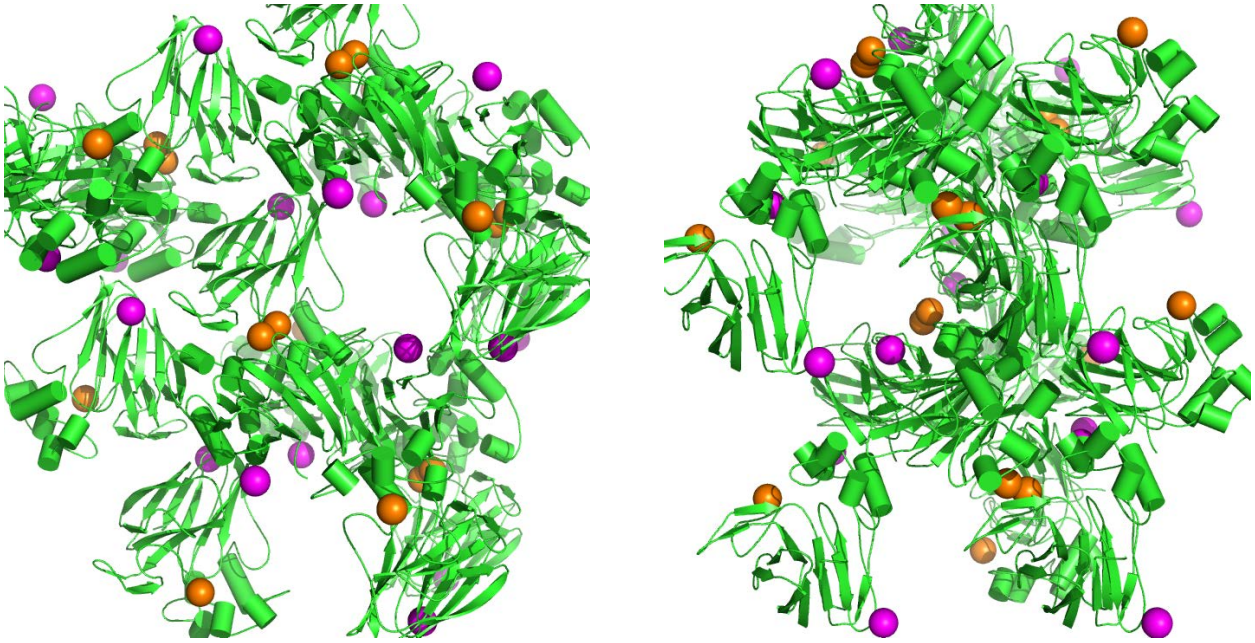


Figure 5.10 – Surface électrostatique du second site de fixation du TbTNTPA. Echelle allant de - 4,0 kT.e^{-1} (charge négative, rouge) à + 4,0 kT.e^{-1} (charge positive, bleu)

5.3.4.3 Le complexe TbTNTPA dans l'empilement cristallin d'un cristal dérivé de TdTHAU

Avoir une vue générale de l'empilement cristallin peut donner une idée de la taille des canaux de solvant, de l'accessibilité aux sites de fixation en cas de trempage et de juger un peu mieux de l'encombrement stérique des sites de fixation du complexe. Comparé à celui du lysozyme (*cf.* chapitre précédent), l'empilement cristallin d'un cristal de thaumatine de *Thaumatococcus daniellii* avec 47 % de solvant dans l'unité asymétrique laisse la place nécessaire à la fixation d'un complexe de lanthanide plus volumineux, tel le TbTNTPA **[Figures** 5.11**]**.



(a) Première orientation **(b)** Seconde orientation, rotation de 90°

Figure 5.11 – Vue générale de l'empilement cristallin d'un cristal dérivé de TdTHAU avec le complexe TbTNTPA. Les molécules de protéines sont en vert, les complexes occupant les sites 1 sont en orange et les complexes occupant les sites 2, en violet.

5.4 Résultats obtenus avec la CaMalDH

On obtient des cristaux dérivés de CaMalDH par trempage dans une solution contenant 60 mM de TbTNTPA et en présence de 40 mM de chlorure de cadmium sans difficulté. Aucune fixation du complexe LnTNTPA à la surface de CaMalDH n'a été détectée. Toutefois, cette expérience a mis en évidence la stabilité du LnTNTPA dans une solution contenant 40 mM d'ions divalents. Dans ces conditions de cristallisation, les autres complexes de la grande famille des picolinates auraient été déstabilisés et nous aurions observé la formation de cristaux de sels de complexe.

5.5 Conclusion

Le LnTNTPA est un complexe qui peut être utilisé dans n'importe quelle condition. En effet, grâce à sa cage, c'est le seul complexe de la grande famille des complexes picolinates qui s'affranchit du problème

de l'influence des ions divalents. Plus luminescent dans le visible que le tris-dipicolinate ou que les tris-triazoledipicolinates de lanthanide, neutre et très soluble, il se manipule aussi bien en trempage qu'en co-cristallisation. Cependant, le complexe LnTNTPA ne s'est pas fixé sur le lysozyme de blanc d'œuf de poule, protéine modèle sur laquelle se sont fixés les autres membres de la grande famille des complexes picolinates. Cela montre que nous avons peut être atteint la taille limite de complexes utilisables dans des cristaux où le pourcentage de solvant dans l'unité asymétrique est inférieur à 45 %.

D'autre part, l'étude menée sur les complexes de la famille des $[Ln(TDPA)_3]^x$ avait mis en évidence que c'est la charge globale des complexes qui induit l'effet supramoléculaire qui conduit alors aux "mauvais cristaux". Nous avions aussi observé que les complexes neutres de cette famille ne se fixent pas à la TdTHAU car cette protéine ne propose que des interactions de type liaison hydrogène. Le fait que le complexe LnTNTPA neutre ait trouvé un autre site de fixation sur la protéine TdTHAU et aucun sur le lysozyme tend à prouver que ce complexe ne peut induire d'effet supramoléculaire.

Conclusions à l'issue de cette étude sur les complexes picolinates luminescents

C'est à la suite des travaux de recherche de Guillaume Pompidor que nous avons voulu vérifier si le tris-dipicolinate de lanthanide pouvait être utilisé en tant qu'agent de nucléation. En effet, seul complexe pour lequel il avait été mis en évidence des propriétés supramoléculaires conduisant à de nouvelles formes cristallines, il semblait prometteur. J'ai donc soumis le complexe $[Ln(DPA)_3]^{3-}$ à une expérience de co-cristallisation automatisée avec six protéines éclectiques et 576 conditions de cristallisation différentes. A l'issue de cette expérience, nous n'avons jamais constaté que la présence du complexe favorisait la nucléation et, plus généralement, nous n'avons pas constaté l'apparition de cristaux dérivés. De plus, l'ensemble de nos observations à l'issue de cette expérience confirme que le complexe $[Ln(DPA)_3]^{3-}$ est inutilisable dans les conditions de cristallisation contenant des ions divalents, des hautes concentrations en sel ou du 2-methyl-2,4-pentanediol. Le complexe $[Ln(DPA)_3]^{3-}$ est donc un complexe qui doit être utilisé avec circonspection en ayant connaissance, au préalable, des conditions de cristallisation de la protéine.

Nous avons participé au développement d'une nouvelle famille de complexes $[Ln(TDPA)_3]^{x}$, dérivés chimiques du $[Ln(DPA)_3]^{3-}$. Ces nouveaux complexes ont permis l'obtention de phases très précises qui ont permis la résolution de la structure du lysozyme de blanc d'œuf de poule et de la thaumatine de *Thaumatococcus daniellii.* Leur affinité pour la surface de ces protéines est bien plus élevée que celle du complexe $[Ln(DPA)_3]^{3-}$ et ces complexes conservent la propriété d'être aisément détectables dans les cristaux de par la luminescence qu'ils émettent dans le visible lorsqu'ils sont excités sous UV proches. Les expériences menées avec ces complexes ont aussi ouvert la voie des complexes dipicolinates hybrides dont l'étude a permis d'entrevoir l'importance de la charge globale pour l'effet supramoléculaire. Cependant, tout comme le complexe $[Ln(DPA)_3]^{3-}$, les complexes de la famille des $[Ln(TDPA)_3]^{x}$ ne sont pas utilisables dans des conditions de cristallisation contenant des ions divalents ou de fortes concentration en sels.

C'est en tenant compte de cette dernière propriété que le nouveau complexe LnTNTPA, dérivé lui aussi du $[Ln(DPA)_3]^{3-}$, a été synthétisé. Ce complexe cagé s'affranchit de l'influence des ions divalents et conserve la propriété d'être luminescent dans le visible. L'utilisation de ce complexe a permis de résoudre la structure de la thaumatine de *Thaumatococcus daniellii* suite à un phasage *de novo.*

Troisième partie

Les complexes de lanthanide pour le phasage de protéines multimériques de grande taille

La seconde partie de cette thèse contient l'ensemble des résultats obtenus suite à l'utilisation des complexes de lanthanide pour le phasage de macromolécules multimériques de grande taille. Cette voie fut ouverte par la résolution de deux structures : celle de l'enzyme dodécamérique chimérique OTCase 3630 de 450 kDa à l'aide du complexe GdHPDO3A (Girard *et al.*, 2003) et celle de Pab87, hydrolase octamérique de 400 kDa (Delfosse *et al.*, 2009). Dans ce dernier cas, le complexe LuHPDO3A permit le phasage de l'octamère entier (groupe d'espace P*1, cf.* introduction). Il n'est donc pas question dans cette partie de "grands complexes protéiques" au sens large mais bel et bien de "grandes protéines multimériques". L'appellation "grands complexes protéiques" pouvant tout aussi bien inclure, par exemple, les complexes hétéroprotéiques d'initiation de transcription ou de traduction.

Prouver que l'utilisation des complexes de lanthanide pour le phasage de grands assemblages protéiques est réalisable émane de deux postulats simples. Le premier postulat est que si les complexes de lanthanide ont trouvé des sites de fixation sur de petites protéines telles que le lysozyme de blanc d'œuf de poule ou la Thaumatine de *Thaumatococcus daniellii*, la surface des protéines de bien plus grande taille devrait proposer davantage de sites de fixation. Le second postulat est lié à l'expression du rapport de Bijvoet telle que décrite dans l'introduction. Il prouve en effet que, pour des taux d'occupation de diffuseurs anomaux identiques et pour des données de diffraction enregistrées à leur seuil d'absorption, un site totalement occupé par un atome de lanthanide permettra de déterminer la structure d'une protéine de masse moléculaire neuf fois plus importante qu'un site totalement occupé par un atome de sélénium (*cf. introduction*).

Cette seconde partie est ainsi constituée de trois chapitres :

Le premier chapitre concerne le phasage de l'amino-peptidase dodécamérique TET1-12s de 480 kDa, grand assemblage protéique issu de l'archée *Pyrococcus Horikoshii*. La réussite de ce phasage illustre le pouvoir phasant des lanthanides et ce, même sur les données à basse résolution.

Le second chapitre est davantage méthodologique. En premier lieu, je présenterai la nouvelle approche mise en place pour approfondir notre connaissance des modes de fixation des différents complexes étudiés. Cette approche s'appuie sur la grande famille des malate déshydrogénases ("MalDH"). En second lieu, je décrirai les trois nouvelles structures de malate déshydrogénases déterminées à haute résolution grâce aux complexes de lanthanide issus de l'imagerie médicale. Je présenterai ainsi les nouveaux éclairages sur l'adaptation halophile que nous avons pu dégager de ces structures.

Enfin, le troisième chapitre est relatif à l'étude de la fixation des complexes de lanthanide menée sur un grand assemblage protéique membranaire : la pompe à efflux membranaire "Acriflavine resistance protein B" de *Escherichia coli*. Cette étude s'est malheureusement révélée infructueuse.

Je profite de cette fin d'introduction pour raconter une anecdote relative aux travaux présentés dans ces chapitres. Au cours de ma première année de thèse, j'ai parcouru les différents laboratoires de l'institut à la recherche de protéines de type membranaire ou grand complexe multimérique afin de disposer de modèles potentiels pour l'étude du mode de fixation de nos complexes de lanthanide. Cela m'a conduit a établir diverses collaborations, notamment celles avec le Dr. Dominique Madern et le Dr. Bruno Franzetti, tous deux membres du Laboratoire de Biophysique Moléculaire à l'époque. Les premiers résultats obtenus avec la malate déshydrogénase de *Chloroflexus aurantiacus* confiée par le Dr. Madern et ceux obtenus avec la protéase TET1 du Dr. Franzetti s'avérèrent extrêmement concluants. Les échange de bons procédés continuant d'aller bon train, ils furent, entre autre, à l'origine de la

fusion de nos deux équipes, en 2010. La nouvelle équipe ainsi crée se nomme Extremophiles and Large Molecular Assemblies group ou E.L.M.A.

Chapitre 6

L'amino-peptidase dodécamérique TET1-12s de *Pyrococcus horikoshii*

La protéine TET1 est une protéase géante auto-compartimentée. De part sa taille, cette amino-peptidase, isolée sous sa forme homododécamérique de 444 kDa au sein du groupe E.L.M.A, s'est avéré une candidate de choix pour illustrer le pouvoir phasant des complexes de lanthanide.

Nous verrons dans ce chapitre que cette amino-peptidase illustre, d'une part, le pouvoir phasant des lanthanides à partir de données à basse résolution et, d'autre part, les capacités supramoléculaire du complexe tris-dipicolinate d'europium. M'appuyant sur la publication ajoutée à la fin de ce chapitre, je ne ferai ici qu'un résumé de cette étude.

6.1 L'amino-peptidase tétraédrique dodécamérique PhTET1-12s

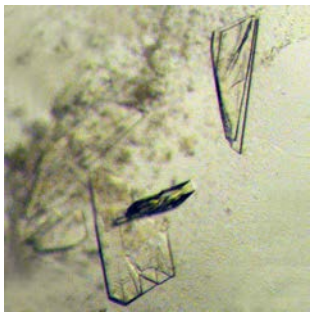
L'amino-peptidase TET a initialement été isolée chez l'archée *Haloarcula marismortui* (Franzetti *et al.*, 2002). Dans le génome de l'archée *Pyrococcus horikoshii*, trois gènes distincts codant des protéines homologues de HmTET ont été identifiés. Elles ont été nommées PhTET1, 2 et 3. Leurs structures tridimentionnelles respectives ont été déterminées (Russo and Baumann, 2004; Borissenko and Groll, 2005; Schoehn *et al.*, 2006). Auto-compartimentées, ces protéines forme des assemblages à douze sous-unités ou 24 sous-unités, nommés TET-12s et TET-24s respectivement (Schoehn *et al.*, 2006). Au sein de la cellule, elles participent à la dégradation de polypeptides qui sont, dans le cas de PhTET1-12s, hydrophiles et de petite taille. La forme octaédrique icositetraedrique ("PhTET-24s") permettrait de dégrader des polypeptides de taille supérieure.

L'analyse de la séquence primaire de la protéine PhTET1-12s conduit à un point isoélectrique théorique de 5,27 et à un coefficient d'extinction molaire de $\varepsilon_{280\,nm} = 26025\ \mathrm{M^{-1}.cm^{-1}}$. La protéine m'a été fournie purifiée et concentrée à 6,2 $\mathrm{mg.mL^{-1}}$ par la Dr. Asunción Durá, membre de l'équipe (Durá *et al.*, 2005).

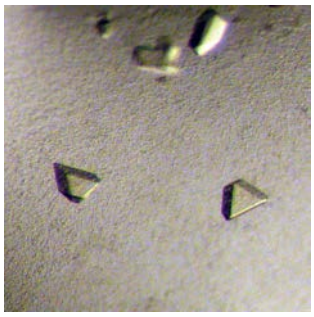
6.2 Obtentions de cristaux natifs et dérivés

La protéine PhTET1-12s peut cristalliser sous deux formes. La première forme, de groupe d'espace $F4_132$, a permis la résolution initiale de la structure à 3,1 Å de résolution (Schoehn *et al.*, 2006). Plus récemment, une nouvelle forme cristalline de symétrie monoclinique (groupe d'espace $P2_1$) a été obtenue.

Ces cristaux, difficilement reproductibles, diffractent à 2,1 Å de résolution. Sur la base des conditions de cristallisation conduisant à cette forme monoclinique, nous avons introduit le complexe tris-dipicolinate d'europium par co-cristallisation. Comme le montre la **figure 6.1**, les cristaux obtenus ne ressemblaient plus aux cristaux natifs.



(a) Cristaux natifs de PhTET1-12s, 560x180x15 µm



(b) Cristaux dérivés de PhTET1-12s, 120x90x60 µm

20- 22 % PEG 3350 / 20 % PEG 2000 MME
100 mM Tris HCl pH 7,5
200 mM TMAO
[P] = 6,2 mg.mL^{-1}
Goutte : 1,5 µL (P) + 1,5 µL (R)
T = 293 K (20°C)
20 - 25 jours
Cryo-protection : 20 % ethylene glycol

(c) Conditions de cristallisation et de cryo-protection

Figure 6.1 – Cristallisation et cryo-protection de PhTET1-12s

6.3 L'effet supramoléculaire et le pouvoir phasant du complexe $[Eu(DPA)_3]^{3-}$

Des données de diffraction ont été enregistrées à partir des cristaux obtenus en présence de $[Eu(DPA)_3]^{3-}$ à l'E.S.R.F et sur la ligne de lumière FIP-BM30A. La longueur d'onde a été choisie au seuil d'absorption L_{III} de l'europium pour que la valeur de la partie imaginaire de son facteur de diffusion atomique soit maximale.

De manière inattendue, l'intégration des données indiqua le groupe d'espace $F4_132$. Ainsi, malgré l'utilisation des conditions de cristallisation conduisant normalement aux cristaux de symétrie monocliniques, la présence du complexe $[Eu(DPA)_3]^{3-}$ a provoqué l'apparition des cristaux de symétrie cubique. Malheureusement, les cristaux ne diffractaient qu'à 4 Å de résolution. Toutefois, les indicateurs anomaux confirmant la fixation du complexe, nous avons décidé de tenter de déterminer la structure sur la base des données SAD disponibles. Un site de fixation par monomère a été déterminé et, malgré la résolution de 4 Å, le phasage a conduit à une carte de densité électronique expérimentale tout à fait interprétable. Comme le montre la **figure 6.2**, la forme générale de la particule est facilement reconnaissable et ce, sous deux orientations caractéristiques.

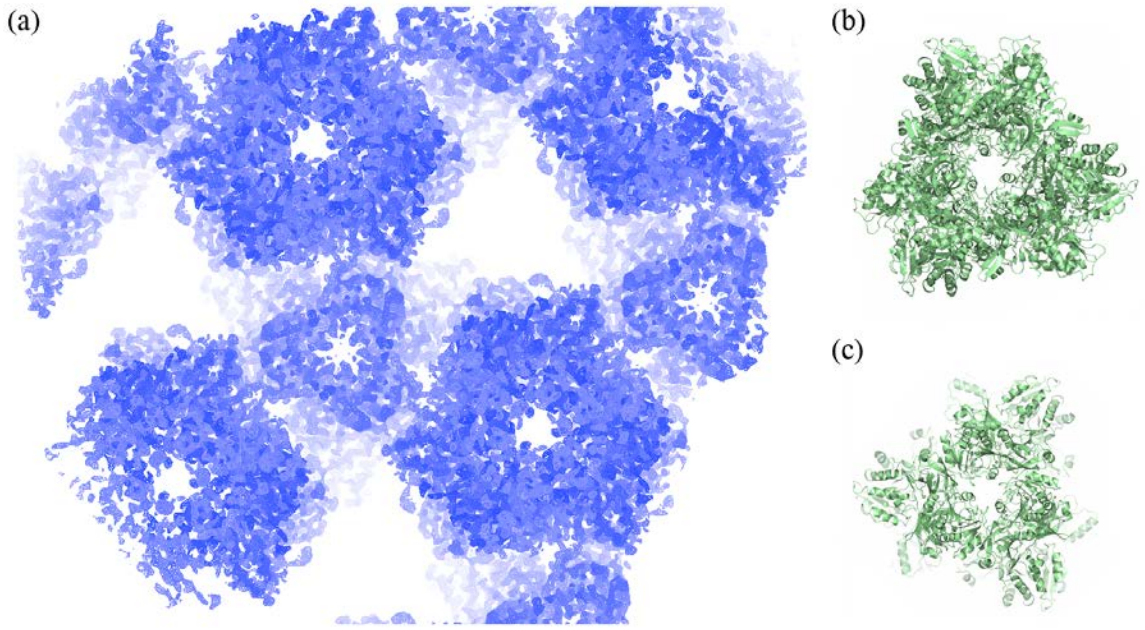
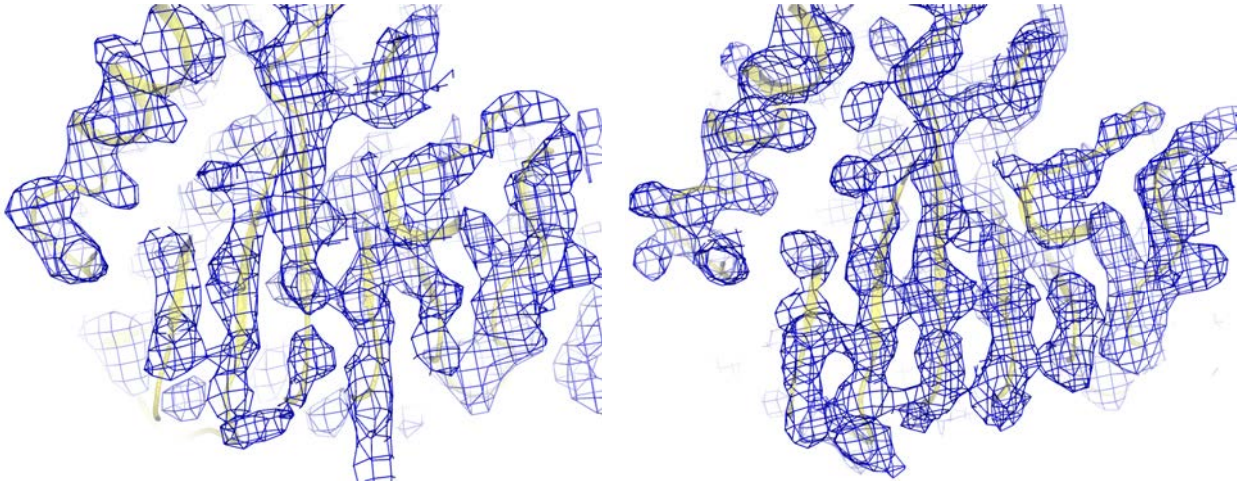


Figure 6.2 – (a) Carte de densité électronique expérimentale SAD (contourée à 1,5 σ) à 4 Å de résolution. (b) Un des pores larges responsables de l'accès des polypetides à l'intérieur de la particule. (c) Un des petits orifices situés sur chacun des quatre sommets

La carte expérimentale à 4 Å de résolution, calculée à partir des données SAD, a permis de deviner aisément la chaine polypeptidique **[figure 6.3]**. Nous avons aussi utilisé les données à 3,1 Å de résolution enregistrées à partir d'un cristal natif pour réaliser un phasage par la méthode SIRAS, celles qui avaient aussi permis de déterminer la structure initiale par remplacement moléculaire (Schoehn *et al.*, 2006). Ce phasage a conduit à une carte expérimentale d'excellente qualité qui indique que l'introduction du complexe $[Eu(DPA)_3]^{3-}$ n'introduit pas un défaut d'isomorphisme **[Figure 6.3]**.



(a) **(b)**

Figure 6.3 – Cartes de densité électronique expérimentales (coutourées à 1,5 σ) obtenues **(a)** par phasage SAD à 4 Å de résolution et **(b)** par phasage SIRAS à 3,1 Å. Un modèle de PhTET1-12s est superposé

6.4 Conclusion

Le complexe tris-dipicolinate d'europium a permis de résoudre la structure de l'amino-peptidase géante TET1-12s. En partant des conditions de cristallisation qui, habituellement, conduisent à des cristaux de symétrie monoclinique, l'introduction du complexe $[Eu(DPA)_3]^{3-}$ a conduit à l'apparition de cristaux de symétrie cubique. En établissant des interactions hydrogènes et hydrophobes (de type empilement-π) avec plusieurs molécules dans le cristal (interactions supramoléculaires), le complexe $[Eu(DPA)_3]^{3-}$peut provoquer un changement de groupe d'espace tel que décrit dans la première partie de ce manuscrit. Toutefois dans le cas des cristaux dérivés de PhTET1-12s, il semble que l'origine du changement de groupe d'espace ne soit pas due à de telles interactions. En effet, bien que la résolution ne permette pas la modélisation complète du complexe, on peut voir que l'ion Eu^{3+} est situé entre deux monomères à proximité du pore large de la particule. Or, ces deux monomères forment le dimère supposé être l'unité de base de la particule entière. Ainsi, le complexe étant lié à l'intérieur de ce bloc de base, ne peut pas influencer directement l'empilement cristallin. Une explication possible du changement de groupe d'espace est que la fixation du complexe $[Eu(DPA)_3]^{3-}$ induit un léger changement dans la conformation globale du protomère de PhTET1-12s qui conduit à la croissance des cristaux cubiques.

De plus, la qualité de la carte de densité électronique obtenue permet de situer le complexe non loin d'une boucle que l'on suppose indispensable pour l'adressage du substrat vers la chambre catalytique de la particule TET (Durá *et al.*, 2009). Ainsi, dans le but d'obtenir de nouvelles information structurales quant à cette zone d'importance, il est prévu d'essayer d'obtenir des données de diffraction à une résolution supérieure. Pour cela, on peut tenter d'obtenir des cristaux dérivés par trempage dans une solution concentrée en complexe $[Eu(DPA)_3]^{3-}$ ou obtenir des cristaux dérivés de PhTET2 ou PhTET3 qui pourraient conduire à de nouvelles informations expérimentales.

electronic reprint

Journal of
Synchrotron Radiation

ISSN 0909-0495

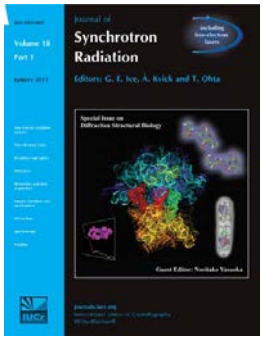
Editors: **G. Ice, Å. Kvick and T. Ohta**

Using lanthanoid complexes to phase large macromolecular assemblies

Romain Talon, Richard Kahn, M. Asunción Durá, Olivier Maury, Frédéric M. D. Vellieux, Bruno Franzetti and Eric Girard

J. Synchrotron Rad. (2011). **18**, 74–78

This open-access article is distributed under the terms of the Creative Commons Attribution Licence http://creativecommons.org/licenses/by/2.0/uk/legalcode, which permits unrestricted use, distribution, and reproduction in any medium, provided the original authors and source are cited.



Synchrotron radiation research is rapidly expanding with many new sources of radiation being created globally. Synchrotron radiation plays a leading role in pure science and in emerging technologies. The *Journal of Synchrotron Radiation* provides comprehensive coverage of the entire field of synchrotron radiation research including instrumentation, theory, computing and scientific applications in areas such as biology, nanoscience and materials science. Rapid publication ensures an up-to-date information resource for scientists and engineers in the field.

Crystallography Journals Online is available from **journals.iucr.org**

Journal of
Synchrotron
Radiation

ISSN 0909-0495

Received 26 May 2010
Accepted 14 September 2010

Using lanthanoid complexes to phase large macromolecular assemblies

Romain Talon,[a,b,c] Richard Kahn,[a,b,c] M. Asunción Durá,[a,b,c] Olivier Maury,[d] Frédéric M. D. Vellieux,[a,b,c] Bruno Franzetti[a,b,c]* and Eric Girard[a,b,c]*

[a]CEA, IBS, F-38054 Grenoble, France, [b]CNRS, UMR5075, F-38027 Grenoble, France, [c]Université Joseph Fourier, F-38000 Grenoble, France, and [d]Université de Lyon, Ecole Normale Supérieure de Lyon, Laboratoire de Chimie, UMR CNRS 5182, 46 Allée d'Italie, 69364 Lyon Cedex 07, France. E-mail: bruno.franzetti@ibs.fr, eric.girard@ibs.fr

Lanthanoid ions exhibit extremely large anomalous X-ray scattering at their L_{III} absorption edge. They are thus well suited for anomalous diffraction experiments. A novel class of lanthanoid complexes has been developed that combines the physical properties of lanthanoid atoms with functional chemical groups that allow non-covalent binding to proteins. Two structures of large multimeric proteins have already been determined by using such complexes. Here the use of the luminescent europium tris-dipicolinate complex $[Eu(DPA)_3]^{3-}$ to solve the low-resolution structure of a 444 kDa homo-dodecameric aminopeptidase, called PhTET1-12s from the archaea *Pyrococcus horikoshii*, is reported. Surprisingly, considering the low resolution of the data, the experimental electron density map is very well defined. Experimental phases obtained by using the lanthanoid complex lead to maps displaying particular structural features usually observed in higher-resolution maps. Such complexes open a new way for solving the structure of large molecular assemblies, even with low-resolution data.

Keywords: lanthanoid complexes; experimental phasing; large macromolecular assemblies; TET aminopeptidase.

1. Introduction

Even though most of the newly deposited structures in the Protein Data Bank (PDB) were solved by molecular replacement, experimental phasing remains essential for determining three-dimensional protein structures if only for solving structures with new folds or which significantly differ from any known model structure. Over the last ten years, methods based on anomalous scattering, namely the single-wavelength anomalous diffraction (SAD) and multiple-wavelength anomalous diffraction (MAD) methods, have replaced the traditional methods based on isomorphous replacement, thus becoming the methods of choice for solving *de novo* protein structures. Consequently, the preparation of effective heavy-atom derivatives displaying anomalous scattering has become a key point for *de novo* crystal structure determination. With the incorporation of selenium through the substitution of methionine residues by seleno-methionine (Hendrickson *et al.*, 1990; Doublie, 1997), and with the developments at third-generation synchrotron radiation sources, which allow weak anomalous signals from intrinsic scatterers to be used, the time-consuming preparation of heavy-atom derivatives has been facilitated.

However, the use of such procedures is not always possible, which revives the problem of incorporating effective anomalous scatterers into protein crystals. Therefore, we proposed to use lanthanoid complexes for preparing lanthanoid derivative crystals (Girard *et al.*, 2002). Lanthanoid ions, Ln^{3+}, are well suited to anomalous diffraction experiments since they all exhibit a strong white line in their L_{III} absorption edge leading to extremely large anomalous contributions of almost 30 e^- for both f' and f''.

A way to assess the phasing power of lanthanides is to compare them with the most frequently used anomalous scatterer, *i.e.* selenium from seleno-methionine. For this purpose the Bijvoet ratio can be considered. We have shown (Girard, Stelter *et al.*, 2003) that the Bijvoet ratio can be expressed as

$$\frac{\Delta F^{\pm}}{F} = \frac{1}{Z_{\rm eff}}\left[\frac{2\sum_j q_j^2 \left(f_j''\right)^2}{N_{\rm P}}\right]^{1/2},$$

where $N_{\rm P}$ is the number of atoms of the protein of mean scattering factor $Z_{\rm eff}$, q_j and f_j'' are the site occupancy and the

electronic reprint

imaginary part of the atomic scattering factor of the anomalous scatterer j, respectively. This formula clearly shows that, assuming fixed site occupancies, identical Bijvoet ratios are obtained for a protein that is four times larger when f_j'' is doubled for each anomalous scatterer. Assuming that the diffraction data are collected at the respective absorption edge, f_j'' values are about 10 and 30 e^- for selenium and lanthanoid, respectively. This means that one fully occupied lanthanoid atom will allow a protein that is nine times larger, compared with a fully occupied Se atom, to be phased.

Hence, lanthanoids are good candidates for macromolecular structure determination based on the use of the anomalous signal. Lanthanoid ions were used in early MAD studies on calcium-binding proteins (Kahn *et al.*, 1985; Weis *et al.*, 1991) as they can substitute for Ca^{2+}. Nagem *et al.* (2001) proposed to incorporate lanthanoid salts through the quick cryo-soak method, but soaking crystals in solutions containing lanthanoid salts often damages the crystals owing to the preferred nine-based coordination of lanthanoid ions. To overcome this problem, Purdy *et al.* (2002) proposed to use a covalent linkage between a lanthanoid complex featuring a saturated coordination sphere and the protein of interest through a thio-reactive functionality, and Silvaggi *et al.* (2007) proposed to use a double lanthanoid-binding tag.

Girard *et al.* (2002) proposed to use gadolinium complexes initially used as contrast agents for magnetic resonant imaging to incorporate lanthanoid ions into protein crystals. Seven different gadolinium complexes were studied (Girard, Stelter *et al.*, 2003). These complexes are made of a ligand that surrounds the lanthanoid ions as a cage, thus providing the majority of the coordination sphere of the ion. More recently, we have proposed to use complexes based on dipicolinate (DPA = pyridine-2,6-dicarboxylate) ligands, namely the lanthanoid tris-dipicolinate complex ions $[Ln(DPA)_3]^{3-}$, the Eu and Tb complexes being luminescent (D'Aléo *et al.*, 2007; Pompidor *et al.*, 2008). As previously mentioned, lanthanoid salts often damage protein crystals even at low concentration. The great advantage of using lanthanoid complexes comes from the fact that the interaction of the lanthanoid ions with the protein occurs through the ligand forming the complexes rather than direct interaction. The binding mode of the various used complexes to the protein turned out to depend on the nature of the ligand (Girard, Anelli *et al.*, 2003). These complexes can be introduced in protein crystals either by co-crystallization or soaking and can be used at rather high concentration (50 to 100 m*M*).

The technique of introducing lanthanoid ions into protein crystals by using lanthanoid complexes[1] was successfully used to solve the structure of several proteins (Chaudhuri *et al.*, 2003; de Bono *et al.*, 2005; Hermoso *et al.*, 2005; Jeudy *et al.*, 2005; Márquez *et al.*, 2006; Gras *et al.*, 2007; Delfosse *et al.*, 2009; Molina *et al.*, 2009; Arnoux *et al.*, 2009; Pérez-Dorado, González *et al.*, 2010; Pérez-Dorado, Sanles *et al.*, 2010).

[1] The different lanthanoid complexes can be purchased from NatX-ray (http://www.natx-ray.com/home.html).

The lanthanoid complexes have also been used to solve structures of large macromolecular assemblies. The structure of a chimeric ornithine carbamoyl transferase, OTCase3630, a dodecamer of 450 kDa, was solved by using the SAD method (Girard, Stelter *et al.*, 2003). More recently, the structure of the *Pyrococcus abyssi* Pab87 protein, an archaeal member of a new self-compartmentalizing protease family forming a cubic-shaped octamer of 400 kDa, was determined at 2.2 Å resolution by the SAD method (Delfosse *et al.*, 2009).

Here, we report the use of the tris-dipicolinate complex to obtain experimental phases at low resolution on a large homo-dodecameric enzyme, PhTET1-12s, which is a tetrahedral aminopeptidase belonging to a new family of self-compartmentalized large protease complexes (Franzetti *et al.*, 2002). The TET peptidase was initially isolated from *Haloarcula marismortui* (Franzetti *et al.*, 2002). In the archae *Pyrococcus horikoshii*, three different open reading frames coding for TET-homologous proteins were identified. These were named PhTET1, 2 and 3. Their three-dimensional structures were determined (Franzetti *et al.*, 2002; Russo & Baumann, 2004; Borissenko & Groll, 2005; Schoehn *et al.*, 2006; Durá *et al.*, 2009). It has been shown (Schoehn *et al.*, 2006) that PhTET1 assembles as a tetrahedral dodecameric particle (called PhTET1-12s for the 444 kDa assembly made up of 12 subunits) or as an octahedral tetracosameric edifice (called PhTET1-24s for the 888 kDa assembly made up of 24 subunits).

Since the TET particles are highly symmetrical molecular edifices formed by a single type of subunit, they provide an excellent model for probing the phasing capacity of different lanthanoid complexes. Moreover, the currently available TET crystallographic structures do not permit detailed analyses of the particles interior. The polypeptide trafficking and the processing mechanisms by the TET particles remain therefore unclear. In this paper we show that low-resolution experimental phase obtained with tris-dipicolinate complex can provide novel structural information on the PhTET1-12s complex.

2. Materials and methods

2.1. Materials

The expression of TET1-12s from *P. horikoshii* in *Escherichia coli*, its purification and initial crystallization assays will be described elsewhere (Dura *et al.*, 2010). $Na_3[Eu(DPA)_3].6H_2O$ complex where DPA stands for pyridine-2,6-dicarboxylate (*e.g.* dipicolinate) was prepared using the procedure described by Tancrez *et al.* (2005).

2.2. Crystallization

Crystallization was performed by vapour diffusion using the hanging-drop method at 293 K. Native PhTET1-12s crystals (~200 µm × 140 µm × 20 µm) were grown within three weeks by mixing 1.5 µl of 6.2 mg ml^{-1} protein solution and 1.5 µl of 20–22% PEG 3350 (or 20% PEG 2000 MME), 100 m*M* Tris-HCl buffer at pH 7.5 and 200 m*M* trimethylamine *N*-oxide

electronic reprint

Table 1
Data collection and processing statistics of the $[Eu(DPA)_3]^{3-}$ derivative of TET1-12s.

Space group	$F4_132$
Unit-cell parameter a	219.83 Å
Resolution range (high-resolution shell)	49.16–4.00 Å (4.22–4.00 Å)
No. of unique reflections	4207 (591)
R_{merge}†	8.6% (22.3%)
R_{pim}‡	3.0% (8.7%)
R_{ano}§	6.8% (7.8%)
$I/\sigma(I)$¶	7.5 (3.5)
Completeness	99.9% (97.9%)
Multiplicity	13.7 (13.7)

† $R_{merge} = \sum_h \sum_i |\bar{I}(h) - I_i(h)| / \sum_h \sum_i |I_i(h)|$ where $I_i(h)$ is the ith measurement of reflection h and $\bar{I}(h)$ is the mean measurement of reflection h. ‡ $R_{pim} = \sum_h [1/(N-1)]^{1/2} \sum_i |I_i(h) - \bar{I}(h)| / \sum_h \sum_i I_i(h)$. This indicator, which describes the precision of the averaged measurement, is most relevant (Weiss, 2001). § $R_{ano} = \sum_h |\bar{I}^+(h) - \bar{I}^-(h)| / \sum_h |\bar{I}^+(h) + \bar{I}^-(h)|$ where $\bar{I}^+(h)$ and $\bar{I}^-(h)$ are the mean intensities of a Friedel mate. ¶ $I/\sigma(I)$ is the signal-to-noise ratio for merged intensities.

reservoir solution. PhTET1-12s derivative crystals (~150 µm × 150 µm × 30 µm) were obtained within three weeks by cocrystallization with $[Eu(DPA)_3]^{3-}$ by mixing 1.5 µl of 6.2 mg ml^{-1} protein solution, 1.5 µl of 220 m*M* $Na_3[Eu(DPA)_3].6H_2O$ solution and 1.5 µl of reservoir solution containing 23–24% PEG 3350 (or 23–26% PEG 2000 MME).

Prior to data collection, derivative crystals were cryo-cooled in liquid nitrogen using mother liquor containing 20% ethylene glycol as cryo-protectant.

2.3. Data collection and data processing

SAD data were collected on the FIP-BM30A beamline at the ESRF. Based on a fluorescence scan, the wavelength was chosen at the L_{III} europium absorption edge, and was set to 1.766 Å, which corresponds to the maximum value of f'' (~28 e^-). Diffraction data were integrated using the program *XDS* (Kabsch, 2010*a*,*b*) and the integrated intensities were scaled and merged using the CCP4 programs *SCALA* and *TRUNCATE* (Collaborative Computational Project, Number 4, 1994). A summary of the processing statistics is given in Table 1.

3. Results

3.1. Derivative crystal form

As described in §2, we used crystallization conditions that led to a new high-resolution form of PhTET1-12s in space group $P2_1$ with an entire dodecamer in the asymmetric unit (Dura *et al.*, 2010). Surprisingly, the addition of the tris-dipicolinate complex led to the initial $F4_132$ crystal form diffracting at low resolution, that was used for the initial structure determination of PhTET1-12s at 3.09 Å resolution (Porciero *et al.*, 2005; Schoehn *et al.*, 2006).

3.2. *De novo* structure determination

As shown in Table 1, the high value of R_{ano} clearly indicated the presence of tris-dipicolinate europium complex binding sites, which was then confirmed by the anomalous Patterson map. Despite the low resolution of the data, we attempted *de novo* phasing of the structure of PhTET1-12s. Using the program *SHELXD* (Sheldrick, 2008), we were able to locate one Eu site per TET-monomer. Heavy-atom refinement and initial phasing were performed using the program *SHARP* (La Fortelle & Bricogne, 1997). Phases from *SHARP* were improved by density modification using the CCP4 program *DM* (Cowtan & Main, 1996) assuming a solvent content of 50%.

3.3. Experimental 4.0 Å SAD phasing

Despite the low resolution, the experimental phases were accurate since the figure of merit after *SHARP* and *DM* are 0.369 and 0.731, respectively. The resulting experimental electron density map was of good quality (Fig. 1*a*) since it allowed the polypeptide chain to be traced, as shown in Fig. 2(*a*). The overall shape of the PhTET1-12 subunit particle could be easily recognized with, on one side of the particle, the large channel (Fig. 1*b*) assumed to be the entrance for the peptide substrate and, on the other side, the small channel (Fig. 1*c*) assumed to be the exit pathway for the reaction products, which are individual amino acids.

3.4. Experimental 3.09 Å SIRAS phasing

We performed SIRAS (single isomorphous replacement with anomalous scattering) phasing using the 3.09 Å resolution native data set from which the structure of PhTET1-12s (PDB code 2cf4) was solved by molecular replacement (Rossmann, 1990). As for the SAD phasing, the SIRAS experimental phases were accurate since the figure of merit

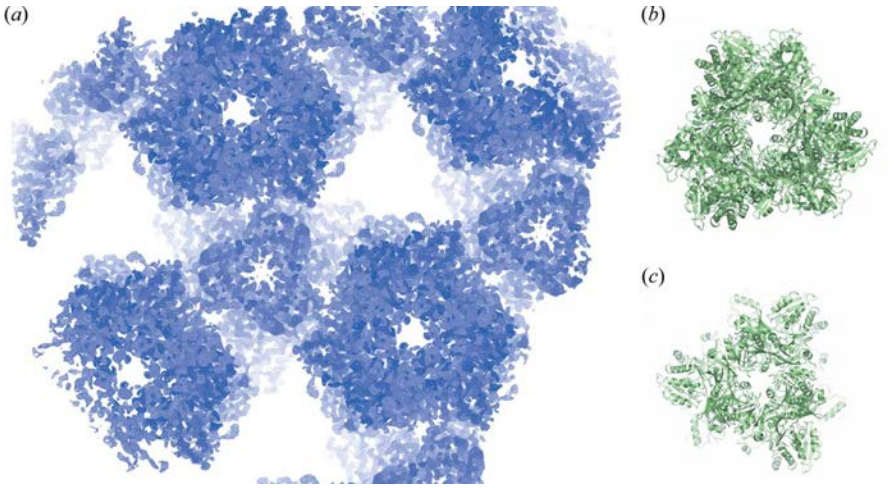


Figure 1
(*a*) Experimental 4.0 Å SAD electron density map contoured at 1.5σ showing (*b*) one of the large pores situated in each facet of the PhTET1-12s particle that gives access to the interior of the system and (*c*) one of the small orifices situated on each of the four apices. Figs. 1 and 2 were prepared using *COOT* (Emsley *et al.*, 2010) and *PyMOL* (http://www.pymol.org/).

electronic reprint

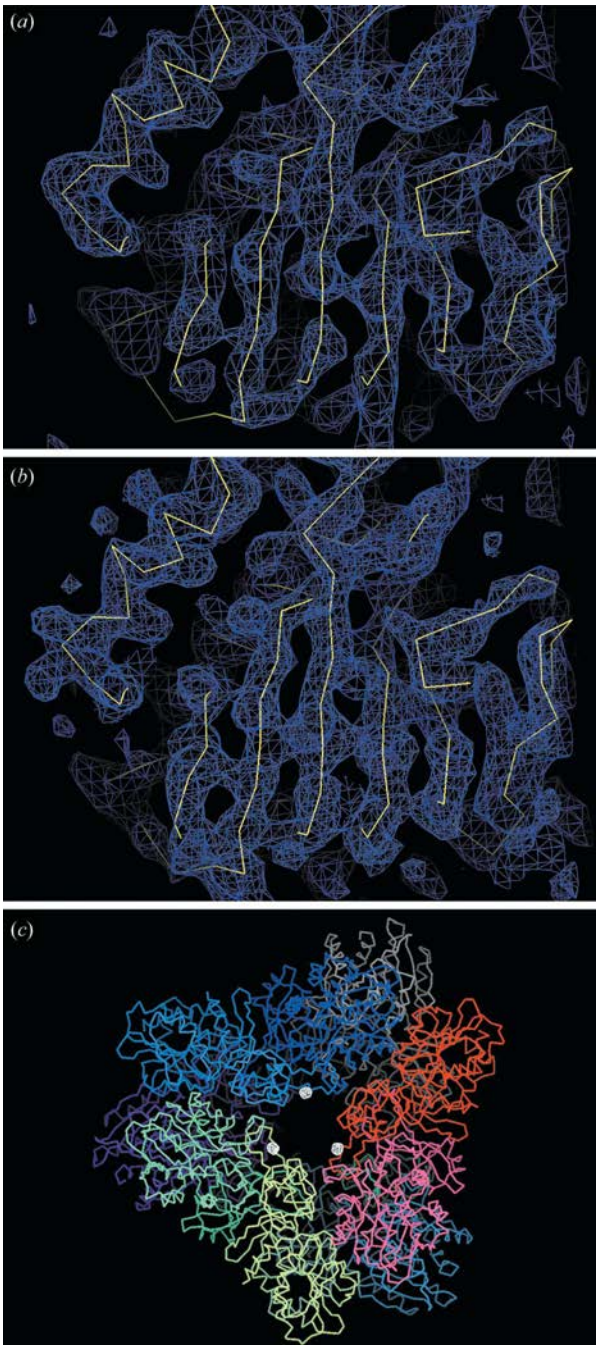


Figure 2
(*a*) Experimental 4.0 Å SAD electron density map. (*b*) Experimental 3.09 Å SIRAS electron density map. Both maps are contoured at 1.5σ. The model shown corresponds to PDB code 2cf4 (Schoehn *et al.*, 2006). (*c*) Anomalous Fourier map contoured at 10σ showing that the Eu^{3+} ion of the $[Eu(DPA)_3]^{3-}$ complex is located between two monomers on the large channel side of the particle. This dimer is considered as the minimal building block to form the whole TET1 particle.

after *SHARP* and *DM* were 0.211 and 0.785, respectively. Despite the introduction of the tris-dipicolinate europium complex, the isomorphism between the native and derivative crystals was preserved. The resulting SIRAS electron density map was of high quality, as shown in Fig. 2(*b*).

4. Conclusion

We have shown that, using $[Eu(DPA)_3]^{3-}$, the high-phasing-power heavy-atom derivative of PhTET1-12s may be obtained by co-crystallization. Highly accurate experimental phases were obtained, even at the low resolution of this work (4.0 Å). The presence of the $[Eu(DPA)_3]^{3-}$ complex modified the crystal space group: from crystallization conditions that led to the monoclinic crystal form diffracting at high resolution, the introduction of $[Eu(DPA)_3]^{3-}$ induced the formation of cubic crystals. Pompidor *et al.* (2010) showed that the interaction between the protein and the $[Ln(DPA)_3]^{3-}$ complex occurs through hydrogen bonds between the O atom of the carboxylate groups of the DPA ligands and hydrogen-bond donor residues, and through hydrophobic π-stacking interaction between DPA rings and aromatic residues. In some cases this specific binding mode improves the protein–protein interaction involved in crystal packing leading to supramolecular interactions. In the present structure it seems that it is not the case. Even if the low resolution of the data limits the modelling of the DPA ligand, the Eu^{3+} ion is located between two monomers on the large channel side of the particle, as shown in Fig. 2(*c*). These two monomers are supposed to be the minimal building block of the whole particle. Since the $[Eu(DPA)_3]^{3-}$ complex is bound within this building block, it did not directly influence the molecular packing as would be the case if bridging two building blocks. A possible explanation for the space group change is that binding of the tris-dipicolinate europium complex induces a small conformational change in the PhTET1-12s protomer, leading to the growth of the low-resolution crystal form.

The tris-dipicolinate europium complex binding site is located in the vicinity of a loop, which is assumed to be a key player in the addressing of the substrate toward the catalytic chambers of the TET particle (Durá *et al.*, 2009). To obtain new insights into this important functional zone, we therefore plan to attempt to increase the resolution of the experimental data either by soaking PhTET1 crystals in solutions containing $[Eu(DPA)_3]^{3-}$ or by preparing $[Eu(DPA)_3]^{3-}$ derivative crystals of PhTET2 or PhTET3, in order to obtain more precise experimental (*i.e.* model-bias free) information.

As mentioned, the binding of the lanthanoid complexes to the protein depends on their ligand, the non-covalent interaction being for example hydrophobic (for the complex Gd-HPDO3A; Girard, Stelter *et al.*, 2003) or through hydrogen bonding between arginine/lysine residues and the dipicolinate complex (Pompidor *et al.*, 2010). Thus, the probability of occurrence of the appropriate binding sites in the protein increases with the protein size. Combined with the strong anomalous signal of the lanthanoid ions, these complexes are thus efficient tools for solving the structure of large macromolecular assemblies, irrespective of their size.

This work was supported by the French National Research Agency (ANR Project MacroTET; BLAN07-3_204002).

electronic reprint

References

Arnoux, P., Morosinotto, T., Saga, G., Bassi, R. & Pignol, D. (2009). *Plant Cell*, **21**, 2036–2044.

Bono, S. de, Riechmann, L., Girard, E., Williams, R. L. & Winter, G. (2005). *Proc. Natl. Acad. Sci. USA*, **102**, 1396–1401.

Borissenko, L. & Groll, M. (2005). *J. Mol. Biol.* **346**, 1207–1219.

Chaudhuri, B. N., Sawaya, M. R., Kim, C. Y., Waldo, G. S., Park, M. S., Terwilliger, T. C. & Yeates, T. O. (2003). *Structure*, **11**, 753–764.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Cowtan, K. D. & Main, P. (1996). *Acta Cryst.* D**52**, 43–48.

D'Aléo, A., Pompidor, G., Elena, B., Vicat, J., Baldeck, P. L., Toupet, L., Kahn, R., Andraud, C. & Maury, O. (2007). *ChemPhysChem*, **8**, 2125–2132.

Delfosse, V., Girard, E., Birck, C., Delmarcelle, M., Delarue, M., Poch, O., Schultz, P. & Mayer, C. (2009). *PLoS ONE*, **4**, e4712.

Doublie, S. (1997). *Methods Enzymol.* **276**, 523.

Durá, M. A. *et al.* (2010). In preparation.

Durá, M. A., Rosenbaum, E., Larabi, A., Gabel, F., Vellieux, F. M. D. & Franzetti, B. (2009). *Mol. Microbiol.* **72**, 26–40.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.

Franzetti, B., Schoehn, G., Hernandez, J.-F., Jaquinod, M., Ruigrok, R. W. H. & Zaccai, G. (2002). *EMBO J.* **21**, 2132–2138.

Girard, É., Anelli, P. L., Vicat, J. & Kahn, R. (2003). *Acta Cryst.* D**59**, 1877–1880.

Girard, É., Chantalat, L., Vicat, J. & Kahn, R. (2002). *Acta Cryst.* D**58**, 1–9.

Girard, É., Stelter, M., Vicat, J. & Kahn, R. (2003). *Acta Cryst.* D**59**, 1914–1922.

Gras, S., Chaumont, V., Fernandez, B., Carpentier, P., Charrier-Savournin, F., Schmitt, S., Pineau, C., Flament, D., Hecker, A., Forterre, P., Armengaud, J. & Housset, D. (2007). *EMBO Rep.* **8**, 569–575.

Hendrickson, W. A., Horton, J. R. & Lemaster, D. M. (1990). *EMBO J.* **9**, 1665–1672.

Hermoso, J. A., Lagartera, L., González, A., Stelter, M., García, P., Martínez-Ripoll, M., García, J. L. & Menéndez, M. (2005). *Nat. Struct. Mol. Biol.* **12**, 533–538.

Jeudy, S., Stelter, M., Coutard, B., Kahn, R. & Abergel, C. (2005). *Acta Cryst.* F**61**, 848–851.

Kabsch, W. (2010*a*). *Acta Cryst.* D**66**, 125–132.

Kabsch, W. (2010*b*). *Acta Cryst.* D**66**, 133–144.

Kahn, R., Fourme, R., Bosshard, R., Chiadmi, M., Risler, J., Dideberg, O. & Wery, J. (1985). *FEBS Lett.* **179**, 133–137.

La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.

Márquez, J., Reinelt, S., Koch, B., Engelmann, R., Hengstenberg, W. & Scheffzek, K. (2006). *J. Biol. Chem.* **281**, 32508–32515.

Molina, R., González, A., Stelter, M., Pérez-Dorado, I., Kahn, R., Morales, M., Moscoso, M., Campuzano, S., Campillo, N. E., Mobashery, S., García, J. L., García, P. & Hermoso, J. A. (2009). *EMBO Rep.* **10**, 246–251.

Nagem, R. A. P., Dauter, Z. & Polikarpov, I. (2001). *Acta Cryst.* D**57**, 996–1002.

Pérez-Dorado, I., González, A., Morales, M., Sanles, R., Striker, W., Vollmer, W., Mobashery, S., García, J. L., Martínez-Ripoll, M., García, P. & Hermoso, J. A. (2010). *Nat. Struct. Mol. Biol.* **17**, 576–581.

Pérez-Dorado, I., Sanles, R., González, A., García, P., García, J. L., Martínez-Ripoll, M. & Hermoso, J. A. (2010). *Acta Cryst.* F**66**, 448–451.

Pompidor, G., D'Aléo, A., Vicat, J., Toupet, L., Giraud, N., Kahn, R. & Maury, O. (2008). *Angew. Chem. Int. Ed.* **47**, 3388–3391.

Pompidor, G., Maury, O., Vicat, J. & Kahn, R. (2010). *Acta Cryst.* D**66**, 762–769.

Porciero, S., Receveur-Bréchot, V., Mori, K., Franzetti, B. & Roussel, A. (2005). *Acta Cryst.* F**61**, 239–242.

Purdy, M. D., Ge, P., Chen, J., Selvin, P. R. & Wiener, M. C. (2002). *Acta Cryst.* D**58**, 1111–1117.

Rossmann, M. G. (1990). *Acta Cryst.* A**46**, 73–82.

Russo, S. & Baumann, U. (2004). *J. Biol. Chem.* **279**, 51275–51281.

Schoehn, G., Vellieux, F. M. D., Asunción Durá, M., Receveur-Bréchot, V., Fabry, C. M. S., Ruigrok, R. W. H., Ebel, C., Roussel, A. & Franzetti, B. (2006). *J. Biol. Chem.* **281**, 36327–36337.

Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.

Silvaggi, N. R., Martin, L. J., Schwalbe, H., Imperiali, B. & Allen, K. N. (2007). *J. Am. Chem. Soc.* **129**, 7114–7120.

Tancrez, N., Feuvrie, C., Ledoux, I., Zyss, J., Toupet, L., Le Bozec, H. & Maury, O. (2005). *J. Am. Chem. Soc.* **127**, 13474–13475.

Weis, W. I., Kahn, R., Fourme, R., Drickamer, K. & Hendrickson, W. A. (1991). *Science*, **254**, 1608–1615.

Weiss, M. S. (2001). *J. Appl. Cryst.* **34**, 130–135.

electronic reprint

Chapitre 7

Les malate déshydrogénases issues d'organismes extrêmophiles

Le mot "collaboration" sous-entend "échange de bons procédés" . C'est dans cette optique que fut établie celle avec le Dr. Dominique Madern, biochimiste spécialiste des enzymes appartenant aux superfamilles des lactate déshydrogénases ("LacDH") et des malate déshydrogénases ("MalDH"), collaboration qui devint assez vite un "travail d'équipe" étant donné la fusion de nos deux équipes pour former le groupe E.L.M.A.

Au sein du groupe E.L.M.A, l'étude de l'adaptation des macromolécules aux environnements extrêmes (hautes pressions, hautes et basses températures, hautes concentrations en sels, hauts ou bas pH, hautes concentrations en métaux, hautes doses de rayonnement) est une des thématiques dont l'analyse des enzymes lactate déshydrogénases et leur paralogues malate déshydrogénases extrêmophiles fait partie intégrante. Ces enzymes sont ubiquitaires et sont donc présentes dans le cytoplasme d'organismes cellulaires dont certains ont la capacité de se développer dans ces environnements dits "extrêmes" (d'où le terme "extrêmophile" donné à ces organismes). Pour s'adapter à de telles conditions, la composition du cytoplasme des cellules de tels organismes doit s'homéostaser avec le milieu extérieur. Ainsi, dans le cas d'organismes vivants dans des milieux à hautes concentrations en sels par exemple (adaptation halophile), le cytoplasme cellulaire contient, lui-aussi, une concentration en un sel (généralement du KCl) supérieure au molaire (Ginzburg and Sachs, 1970). Ainsi, c'est en voulant comprendre comment des partenaires protéiques, impliqués dans des mécanismes cellulaires ubiquitaires essentiels, restent actifs dans de telles conditions de stress que l'équipe s'est intéressée à la famille des enzymes malate déshydrogénases.

En effet, dans le cas des organismes eucaryotes, cette famille d'oxidoreductases (classification enzymatique E.C. 1.1.1.37) est impliquée dans la dernière étape du cycle de Krebs où ces enzymes catalysent l'oxydation réversible du L-Malate en oxaloacétate via l'utilisation du co-facteur NAD^+ (Ying, 2008). En oxydant le L-Malate issu de la mitochondrie, ces enzymes participent non-seulement à l'indispensable gluconéogénèse mais aussi au renouvellement du stock de protons disponibles pour la cellule sous forme de NADH, agent réducteur essentiel à de nombreuses réactions enzymatiques (Ying, 2008).

La séquence primaire des enzymes de cette famille possède normalement une numérotation qui lui est propre (Eventoff *et al.*, 1977). Je précise ici que dans ce chapitre, je n'utiliserai pas cette numérotation mais une numérotation linéaire.

Au sein du groupe E.L.M.A, l'adaptation moléculaire des malate déshydrogénases est étudiée selon trois approches :

- ▷ une approche biochimique qui permet d'étudier en solution, la stabilité et l'activité des MalDH purifiées,
- ▷ une approche phylogénétique rigoureuse qui conduit à déterminer les distances génétiques qui séparent les différentes MalDH de différents taxons au sein d'arbres phylogénétiques. Le calcul de ces arbres s'appuie sur des méthodes heuristiques qui utilisent l'information phylogénomique (concaténation de gènes, ARN ribosomiques), et celle contenues dans les séquences codantes des MalDH,
- ▷ une approche phylostructurale qui permet de replacer l'étude des relations structure-fonction de ces enzymes dans un contexte évolutif. En tenant compte des liens phylogénétiques qui lient les MalDH comparées, on met en lumière les changements structuraux observés entre ces MalDH pour expliquer quels sont ceux qui ont permis telle ou telle adaptation extrêmophile au cours du temps.

7.1 Les malate déshydrogénases pour une étude rationalisée du mode de fixation des complexes de lanthanide

Auprès du Dr. Madern, nous disposons donc d'une "banque de protéines" dont les études déjà menées montrent qu'elles possèdent la même structure quaternaire (homotétramère, 300 à 330 acides aminés par monomère, RMSD $\geq$ 0,95 Å entre les structures alignées) et une similarité de séquence primaire de 72 % **[Figures 7.1 et 7.2]**. Toutefois, malgré ces similitudes, ces protéines ont des propriétés de surface différentes, en lien avec les conditions physico-chimiques auxquelles le cytoplasme de l'organisme dont elles sont issues doit faire face **[Figure 7.3]**. Ainsi, l'idée a germé d'utiliser ces enzymes homotetramériques comme une banque de fonctions chimiques pour comprendre les modes de fixation des complexes de lanthanide. En effet, selon les mutations d'acides aminés liées à l'adaptation extrêmophiles et les positions conservées, on dispose d'un ensemble de conditions physico-chimiques qui nous permettent d'étudier finement les interactions requises pour que les complexes se lient à la protéine. En outre, le fait que les MalDH cristallisent dans au moins deux conditions différentes engendre des variations supplémentaires au niveau de la surface protéique, augmentant ainsi la chimiothèque.

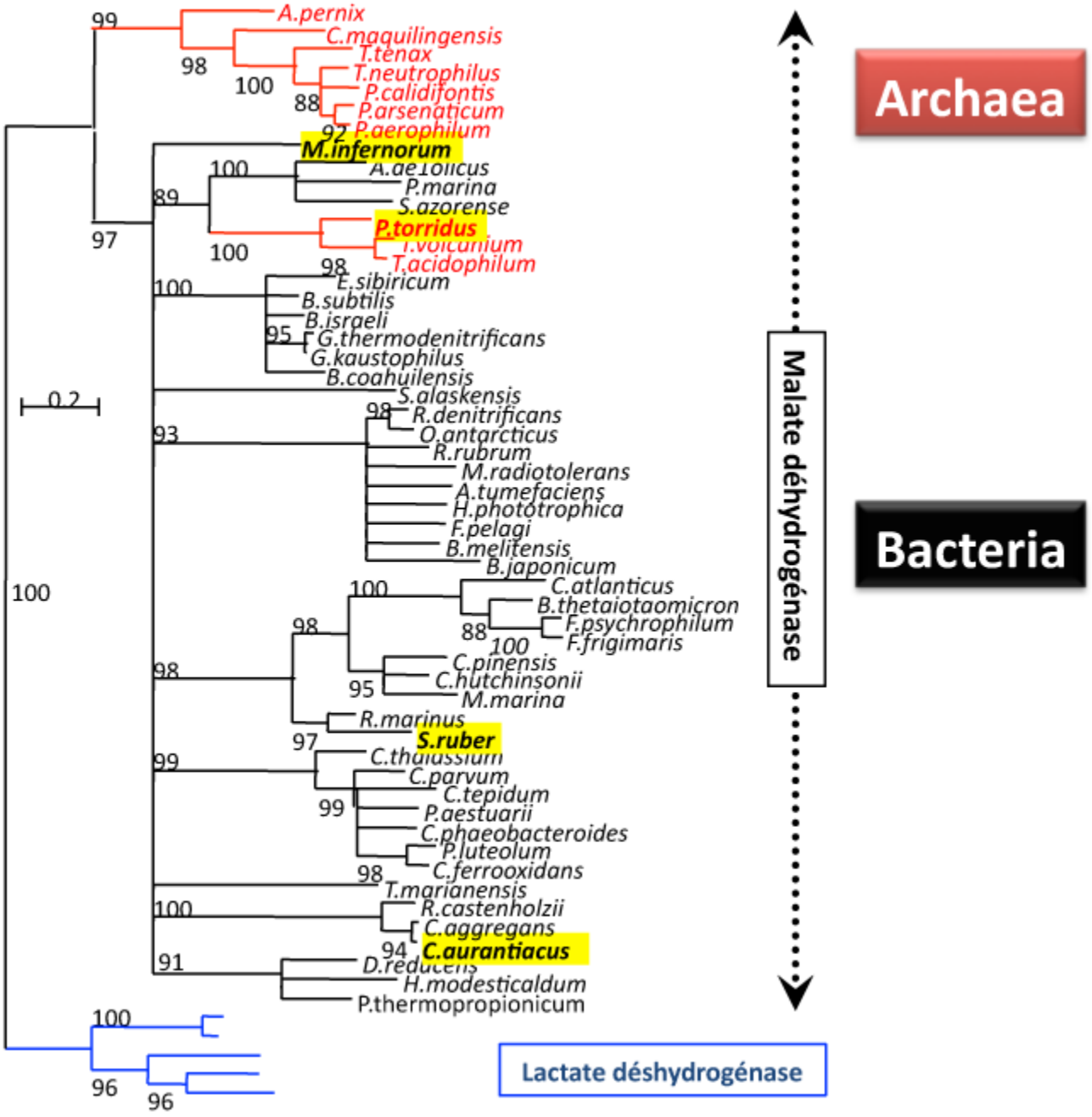


Figure 7.1 – Section d'arbre phylogénétique montrant les distances génétiques entre les différentes MalDH

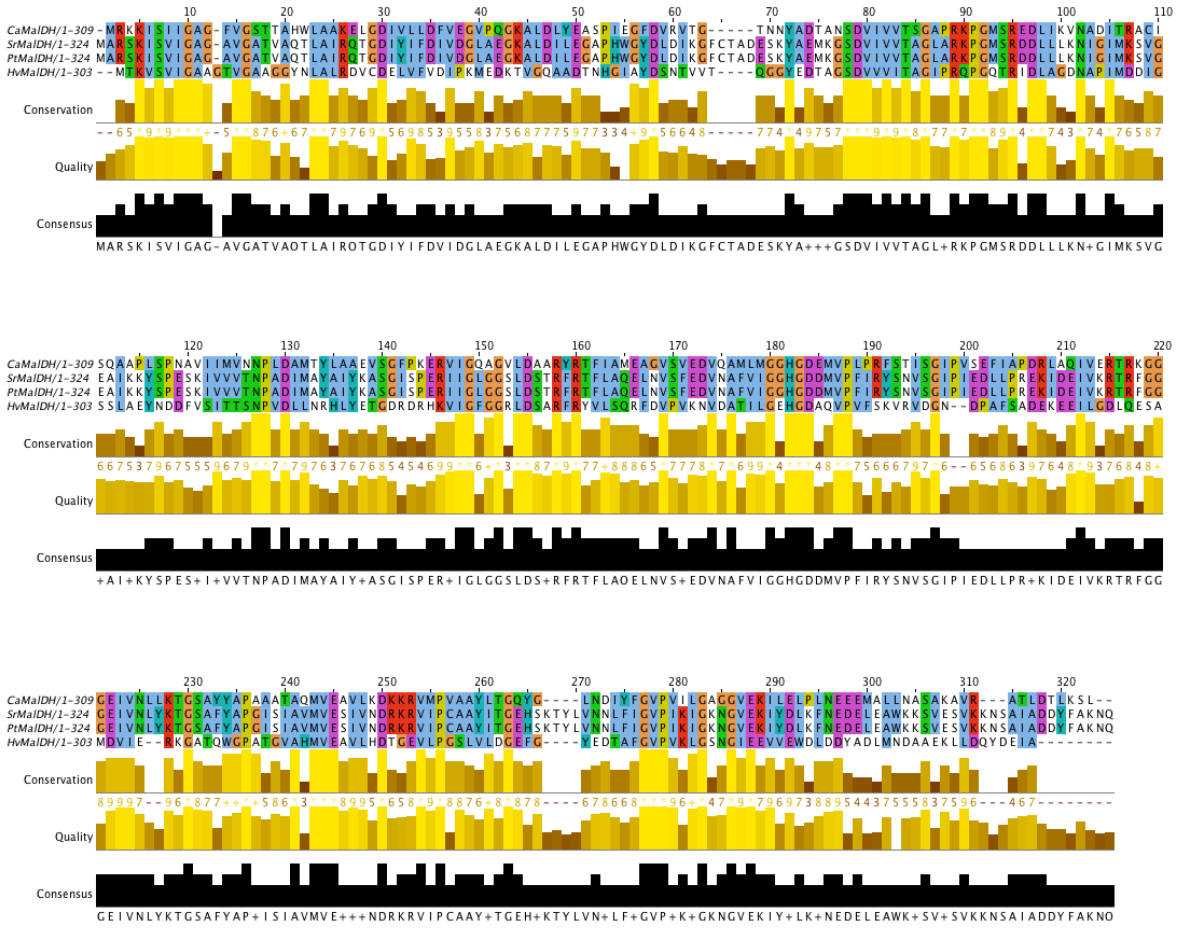


Figure 7.2 – Alignement des séquences primaires des malate déshydrogénases issues des organismes *Chloroflexus aurantiacus* ("CaMalDH"), *Salinibacter ruber* ("SrMalDH"), *Haloferax volcanii* ("HvMalDH") et *Picrophilus torridus* ("PtMalDH"). Alignement réalisé à l'aide du programme Clustal W2, dans sa version 2.1, en utilisant la matrice "Gonnet". Programme disponible en ligne à partir du site internet NCBI. Le rendu provient de l'éditeur d'alignement Jalview version 2.6.1 (Waterhouse *et al.*, 2009).

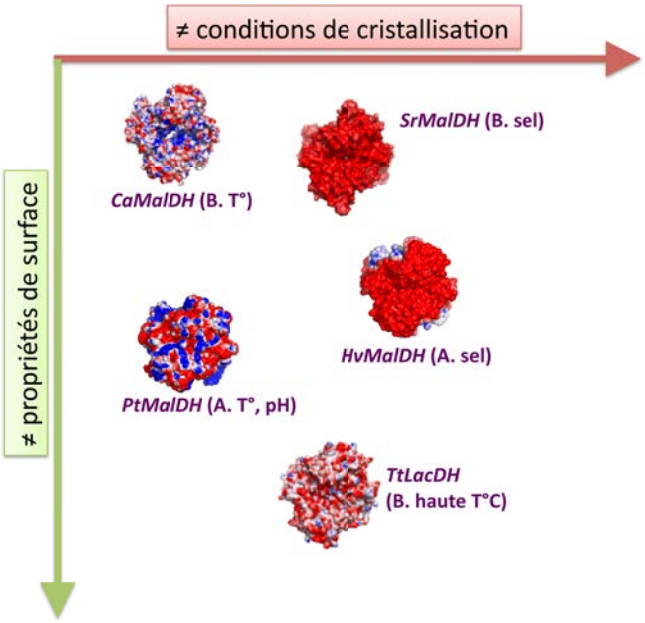


Figure 7.3 – Principe de l'approche rationalisée utilisant les MalDH pour l'étude du mode de fixation des complexes IRM. Surfaces électrostatiques représentées du plus électronégatif (-3 kT/e) au plus électropositif (-3kT/e). CaMalDH pour "MalDH de la bactérie ("B") thermophile ("T°") *Chloroflexus aurantiacus*", SrMalDH pour "MalDH de la bactérie halophile *Salinibacter ruber*", PtMalDH pour "MalDH de l'archée ("A") hyperthermoacidophile ("T°, pH") *Picrophilus torridus*, HvMalDH pour "MalDH de l'archée halophile *Haloferax volcanii* et TtLacDH pour "LacDH de la bactérie hyperthermophile *Thermus thermophilus*".

C'est ici que réside "l'échange de bons procédés". En effet, ces différentes malate déshydrogénases recombinantes nous permettent de mettre en place une méthode d'analyse des modes de fixation plus rationalisée que l'approche menée jusqu'à présent qui consiste à étudier des complexes sur des protéines très différentes. Mais les nouvelles structures résolues à l'aide des complexes sont ensuite confiées en retour au Dr. Madern pour analyses approfondies.

Ainsi je présenterai dans ce chapitre les premiers résultats relatifs à la fixation des complexes IRM GdHPDO3A, GdDO3A et GdDTPA-BMA sur trois MalDH, citées ci-dessous, qui permettent de faire la "preuve de concept" de cette nouvelle approche. Dans les sections suivantes, je présenterai trois nouvelles structures de malate déshydrogénases résolues par phasage expérimental avec les complexes de lanthanide :

1. la structure de la malate déshydrogénase issue de l'archée hyperthermoacidophile *Picrophilus torridus* ("PtMalDH"), résolue grâce au complexe GdHPDO3A,
2. la structure de la malate déshydrogénase issue de l'archée halophile *Haloferax volcanii* ("HvMalDH"), résolue grâce aux complexes GdDO3A et GdDTPA-BMA,
3. la structure de la malate déshydrogénase issue de la bactérie thermophile *Chloroflexus aurantiacus* ("CaMalDH"), résolue grâce aux complexes GdHPDO3A et GdDO3A.

Nous verrons aussi en quoi la construction de ces modèles à haute résolution permet de fournir des informations-clés pour mieux comprendre l'adaptation de ces enzymes au stress auquel elles doivent faire face dans le cytoplasme de la cellule.

7.1.1 Cristallisation des différentes malate déshydrogénases et obtention de cristaux dérivés

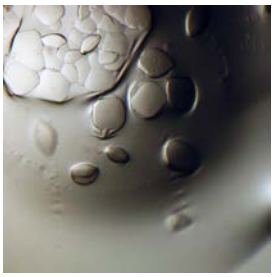
Les PtMalDH, HvMalDH et CaMalDH purifiées et respectivement concentrées à 10 mg.mL^{-1}, 10 mg.mL^{-1} et 20 mg.mL^{-1} m'ont été fournies par le Dr. Madern. Un échantillon concentré à 10 mg.mL^{-1} de chacune d'elles a été amené au robot de cristallisation nanogouttes HTX de l'EMBL. Pour la recherche des conditions de cristallisation natives, 576 conditions de cristallisation fournies par les six kits de cristallisation commerciaux (les mêmes que ceux énoncés dans le premier chapitre de la première partie) ont été testées. Plusieurs conditions de cristallisation ont pu être déterminées pour chaque MalDH.

Si nous avons isolé plusieurs de ces conditions pour obtenir des cristaux natifs de MalDH, nous n'en avons pour l'instant utilisé qu'une seule par MalDH pour obtenir les cristaux dérivés avec les complexes IRM.

Tous les cristaux ont été vitrifiés à 100 K dans l'azote liquide après avoir été trempés dans une solution cryo-protéctante en accord avec leurs conditions de cristallisation.

7.1.1.1 Cristallisation de la malate déshydrogénase de *Picrophilus torridus*

Conditions pour obtenir les cristaux natifs de PtMalDH Deux conditions de cristallisation différentes conduisent à l'apparition de cristaux de PtMalDH de groupe d'espace quadratique **[Figure 7.4]**. La condition avec de hautes concentrations en sulfate d'ammonium permet d'avoir des cristaux de groupe d'espace $I4_122$ avec un dimère dans l'unité asymétrique. La condition en PEG permet d'obtenir des cristaux de groupe d'espace $P4_12_12$ avec un tétramère dans l'unité asymétrique.



(a) Cristaux natifs de PtMalDH obtenus en présence de sulfate d'ammonium, 360x290 µm

1,2 - 1,8 M Sulfate d'ammonium	9 - 14 % PEG 4000
100 mM acide citrique pH 3,5 / 4,0	100 mM acide citrique pH 3,0 / 3,5
[P] = 10 mg.mL^{-1}	
Goutte : 1,5 µL (P) + 1,5 µL (R)	
T = 293 K (20°C)	
8 à 10 jours	12 à 15 jours
Cryo-protection : 30 % glycérol	Cryo-protection : 30 % glycérol

(b) Conditions de cristallisation et de cryo-protection

Figure 7.4 – Cristallisation et cryo-protection de la PtMalDH

Conditions pour obtenir les cristaux dérivés de PtMalDH Les cristaux dérivés de PtMalDH ont été obtenus par co-cristallisation avec 100 mM des complexes GdHPDO3A, GdDO3A et GdDTPA-BMA dans la condition contenant du sulfate d'ammonium **[figure 7.4]**. La cryo-protection des cristaux dérivés a aussi été effectuée avec ajout de 30 % de glycérol aux conditions de la goutte. De l'huile de paraffine a été ajoutée pour le cristal dérivé avec le GdHPDO3A.

7.1.1.2 Cristallisation de la malate déshydrogénase de *Haloferax volcanii*

Conditions pour obtenir les cristaux natifs de HvMalDH La meilleure condition de cristallisation pour obtenir rapidement des cristaux de HvMalDH diffractant à 1,5 Å de résolution est décrite dans la **figure 7.5**. Nous n'avons utilisé que cette condition de cristallisation.

(a) Cristaux natifs de HvMalDH, 240x150 µm

31 - 36 % PEG 400
100 mM Tris HCl pH 8,0
100 mM $MgCl_2$
[P] = 10 $mg.mL^{-1}$
Goutte : 1,5 µL (P) + 1,5 µL (R)
T = 293 K (20°C)
2 à 4 jours
Cryo-protection : la liqueur mère est déjà cryo-protectante

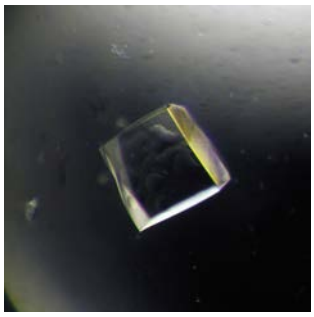
(b) Conditions de cristallisation et de cryo-protection

Figure 7.5 – Cristallisation et cryo-protection de la HvMalDH

Conditions pour obtenir les cristaux dérivés de HvMalDH Les cristaux dérivés de HvMalDH ont été uniquement obtenus par trempage, de 3 secondes à 10 minutes, dans une solution contenant 100 mM de chacun des trois complexes GdHPDO3A, GdDO3A et GdDTPA-BMA en plus des constituants cités dans la **figure 7.5**. L'obtention de cristaux dérivés par co-cristallisation s'est révélée impossible avec la condition de cristallisation utilisée. La cryo-protection des cristaux dérivés est identique à celle des cristaux natifs. Je rappelle ici qu'après chaque trempage, j'ai systématiquement effectué un "back-soak" (*cf.* introduction).

7.1.1.3 Cristallisation de la malate déshydrogénase de *Chloroflexus aurantiacus*

Conditions pour obtenir les cristaux natifs de CaMalDH La condition de cristallisation principalement utilisée est inspirée de celle décrite par B. Dalhus et al, 2002 (Dalhus *et al.*, 2002) et des observations issues de l'expérience de co-cristallisation automatisée avec le complexe $[Eu(DPA)_3]^{3-}$ décrite dans le premier chapitre de la première partie.



(a) Cristal natif de CaMalDH, 800x750 µm

4 - 14 % PEG 400
100 mM acétate de sodium pH 4,6
40 mM acétate de de cadmium
[P] = 20 mg.mL^{-1}
Goutte : 1,5 µL (P) + 1,5 µL (R)
T = 293 K (20°C)
1 à 3 jours
Cryo-protection : 25 % PEG 400

(b) Conditions de cristallisation et de cryo-protection

Figure 7.6 – Cristallisation et cryo-protection de la CaMalDH

Conditions pour obtenir les cristaux dérivés de CaMalDH Les cristaux dérivés de CaMalDH ont été obtenus avec les complexes GdHPDO3A, GdDO3A et GdDTPA-BMA uniquement par trempage. L'obtention de cristaux dérivés par co-cristallisation s'est révélée impossible. Les trempages ont été réalisés dans une solution contenant 100 mM de complexes de lanthanide, en plus des constituants cités dans la **figure 7.6** et d'une durée allant de 3 secondes à 10 minutes. La cryo-protection des cristaux dérivés est identique à celle des cristaux natifs.

7.1.2 L'étude de l'interaction des trois complexes IRM sur les trois MalDH

La comparaison des structures issues du phasage *de novo*, complexe par complexe et site de fixation par site de fixation, permettra de déterminer les caractéristiques des sites protéiques préférentiels pour la fixation de chacun des composés et celles rédhibitoires pour leur fixation. Les structures résolues par remplacement moléculaire, dues à une trop faible fixation des complexes, permettront de moduler et d'affiner les observations précédentes.

Dans un premier temps, le calcul de phases a systématiquement été réalisé par les méthodes de phasage expérimental SAD et MAD. Lorsque le phasage expérimental n'a pu être mené à terme, nous n'avons pas utilisé ici le script utilisé dans la première partie mais le programme de Andrea Thorn et George M. Sheldrick : *ANODE* (Thorn and Sheldrick, 2011) pour détecter la fixation des complexes par synthèse de Fourier anomale. Si des pics élevés dans la synthèse de Fourier anomale ont été détectés, un remplacement moléculaire est effectué (selon la méthode décrite dans l'introduction) pour construire un modèle dans lequel les diffuseurs anomaux sont alors placés. Ceci permet d'avoir une idée des caractéristiques des sites qui ne permettent qu'une faible fixation des complexes IRM. Ensuite, un affinement de ses modèles est réalisé dans le but de déterminer un taux d'occupation pour chacun des sites, y compris pour ceux faiblement occupés. Afin de ne pas alourdir le manuscrit, seuls les tableaux de données relatifs à l'exemple donné en guise de preuve de concept seront reportés dans cette section. Les jeux de données anomales relatifs à PtMalDH et HvMalDH ont été enregistrés au seuil d'absorption L_{III} du gadolinium, lanthanide utilisé pour cette étude. Les jeux de données anomales relatifs à CaMalDH ont été enregistrées au seuil d'absorption Kα du cuivre.

7.1.3 Etude de la fixation du GdHPDO3A sur trois MalDH

Pour illustrer la preuve de concept de cette approche, je présente ici les premiers résultats obtenus suite à une étude cristallographique comparative menée sur les trois MalDH précitées : CaMalDH, PtMalDH et HvMalDH. Suite à l'obtention de cristaux dérivés potentiels, nous avons systématiquement comparé les résultats obtenus avec le complexe GdHPDO3A et les trois MalDH à chacune des étapes menant à la construction et à l'affinement du modèle

7.1.3.1 L'obtention de cristaux dérivés avec le complexe GdHPDO3A

Nous avons déjà constaté des différences au niveau de l'obtention de cristaux dérivés en présence du complexe GdHPDO3A. Introduit en co-cristallisation avec les protéines CaMalDH et HvMalDH, le complexe a empêché toute nucléation. Nous avons pu cependant obtenir des cristaux dérivés de ces protéines par trempage de 3 secondes avec CaMalDH et 10 minutes avec HvMalDH. Seule la protéine PtMalDH a pu être co-cristallisée avec le complexe.

7.1.3.2 Intégration des données de diffraction issues des cristaux dérivés de CaMalDH, PtMalDH et HvMalDH

Le but de cette étude étant l'obtention d'informations précises sur le mode de fixation des complexes, il est essentiel que les données enregistrées soient de qualité élevée. Les valeurs statistiques consignées dans les tableaux **7.1, 7.2 et 7.3** permettent de juger la qualité des données enregistrées sur les cristaux dérivés obtenus en présence de GdHPDO3A.

Intégration des données de diffraction issues d'un cristal de CaMalDH dérivé avec le complexe GdHPDO3A

	GdHPDO3A
	CaMalDH
Méthode	trempage (100 mM de complexe, 3 secondes)
Situation	Institut de Biologie Structurale
Source de rayons X	Générateur de laboratoire
MAD	non
λ (Å)	1,541 ($Cu_{K\alpha}$)
Groupe d'espace	$P3_121$
Paramètres de maille (Å)	a = 106,775, c = 103,526
Contenu unité asymétrique	dimère A-D
Résolution (Å)	19,63 - 1,67 (1,76 - 1,67)
Nombre de réflexions uniques	78541 (10741)
R_{merge} (%)	7,1 (75,5)
R_{pim} (%)	3,4 (40,2)
Sig_{ano} XDS	1,29 (0,85)
R_{ano} (%)	4,3 (28,9)
$\frac{I}{\sigma(I)}$	9,6 (1,0)
Complétude (%)	99,1 (94,0)
Redondance	10,3 (8,2)

Tableau 7.1 – Statistiques d'intégration du jeu de données de diffraction d'un cristal dérivé de CaMalDH avec le complexe GdHPDO3A

J'ai conservé une haute résolution de 1,67 Å malgré une valeur de R_{merge} de 75 %, à posteriori, car j'ai constaté que la haute résolution était essentielle pour entrevoir et/ou modéliser le ligand des complexes.. J'aurais donc pu couper en résolution afin d'obtenir de meilleures statistiques. Toutefois, compte tenu de la forte redondance, la valeur de l'indicateur R_{pim} est de 40 % peut aussi justifier notre choix de conserver une résolution de 1,67 Å.

Intégration des données de diffraction issues d'un cristal de PtMalDH dérivé avec le complexe GdHPDO3A

	GdHPDO3A
	PtMalDH
Méthode	co-cristallisation (100 mM de complexe)
Source synchrotron	SOLEIL
Ligne de lumière	PROXIMA1
MAD	non
λ (Å)	1,711 ($Gd_{L_{III}}$)
Groupe d'espace	$I4_122$
Paramètres de maille (Å)	a = 112,402, c = 286,400
Contenu unité asymétrique	dimère A-D
Résolution (Å)	47,73 - 2,54 (2,68 - 2,54)
Nombre de réflexions uniques	29874 (3601)
$\mathbf{R}_{merge}$ **(%)**	5,8 (29,7)
$\mathbf{R}_{pim}$ **(%)**	3,5 (19,5)
$\mathbf{Sig}_{ano}$ **XDS**	1,43 (0,75)
$\mathbf{R}_{ano}$ **(%)**	4,9 (19,7)
$\frac{I}{\sigma(I)}$	11,4 (2,6)
Complétude (%)	97,3 (81,1)
Redondance	6,7 (4,8)

Tableau 7.2 – Statistiques d'intégration du jeu de données de diffraction d'un cristal dérivé de PtMalDH avec le complexe GdHPDO3A

Intégration des données de diffraction issues d'un cristal de HvMalDH dérivé avec le complexe GdHPDO3A

	GdHPDO3A
	HvMalDH
Méthode	trempage (100 mM de complexe, 10 minutes)
Source synchrotron	ESRF
Ligne de lumière	ID23-1
MAD	non
λ (Å)	1,711 ($Gd_{L_{III}}$)
Groupe d'espace	C2
Paramètres de maille (Å,°)	a = 144,370, b = 82,028, c = 112,325
	β = 101,626
Contenu unité asymétrique	tétramère
Résolution (Å)	48,40 - 1,95 (2,05 - 1,95)
Nombre de réflexions uniques	91862 (12805)
R_{merge} (%)	8,3 (36,4)
R_{pim} (%)	6,4 (27,8)
Sig_{ano} XDS	1,20 (0,70)
R_{ano} (%)	7,3 (21,7)
$\frac{I}{\sigma(I)}$	5,1 (2,1)
Complétude (%)	98,0 (94,1)
Redondance	5,1 (5,0)

Tableau 7.3 – Statistiques d'intégration du jeu de données de diffraction d'un cristal dérivé de HvMalDH avec le complexe GdHPDO3A

Les indicateurs statistiques calculés à partir des données de diffraction enregistrées à partir de ces trois cristaux dérivés montrent la qualité générale de ces dernières. La valeur élevée de l'indicateur R_{ano} pour les trois jeux de données indique une fixation du complexe dans les cristaux de CaMalDH, PtMalDH et HvMalDH.

7.1.3.3 Phasage des données de CaMalDH, PtMalDH et HvMalDH avec le GdHPDO3A

Un phasage expérimental a été possible dans le cas de CaMalDH et PtMalDH. Le phasage des données de CaMalDH a conduit à une carte de densité électronique expérimentale **[Figure 7.7]** facilement interprétable par le programme de construction automatique : 592 acides aminés sur 618 ont été automatiquement construits **[Tableau 7.4]**. Même constat pour PtMalDH ou le phasage *de novo* a conduit à la construction automatique de 634 acides aminés sur 648 **[Tableau 7.5]**. Comme nous l'avons précédemment indiqué, le programme *ANODE* a été utilisé pour confirmer le nombre de sites.

Dans le cas de HvMalDH, le calcul des phases n'a pu se faire par les méthodes de phasage expérimental, nous avons eu donc recours au remplacement moléculaire (selon le protocole indiqué dans l'introduction) **[Tableau 7.6]**.

Phasage expérimental des données CaMalDH avec le complexe GdHPDO3A

	GdHPDO3A
	CaMalDH
Type de données	$SAD_{Cu_{K\alpha}}$
Résolution (Å)	20,61 - 1,67 (1,76 - 1,67)
Pics$_{Fa}$ **(Pic**$_{Soufre}$**) pk (σ)**	37,62 / 31,17 / 30,31 / 22.19 22.16 / 20.69 / 19.85 / 15.41 (12,24)
FOM$^{acen.}_{sharp}$ **/ FOM**$^{cen.}_{sharp}$	0,338 / 0,028
Nombre de sites	8
Résolution FOM$^{acen.}_{sharp} \approx$ **0,5 (Å)**	3,30
P$\varphi^{\neq anomales}_{sharp\,acen.}$ **/ P**$\varphi^{\neq isomorphes}_{sharp\,acen.}$	1,17 / 0,00
SIGMAA$_{solomon}$	0,73
Arp/Warp (construit / total)	592 / 618

Tableau 7.4 – Statistiques de phasage expérimental de CaMalDH avec le complexe GdHPDO3A

La **figure 7.7** met en évidence la précision de la carte expérimentale, améliorée par aplatissement de solvant, qui a permis la construction automatique de 95,8 % du dimère A-D, contenu dans l'unité asymétrique.

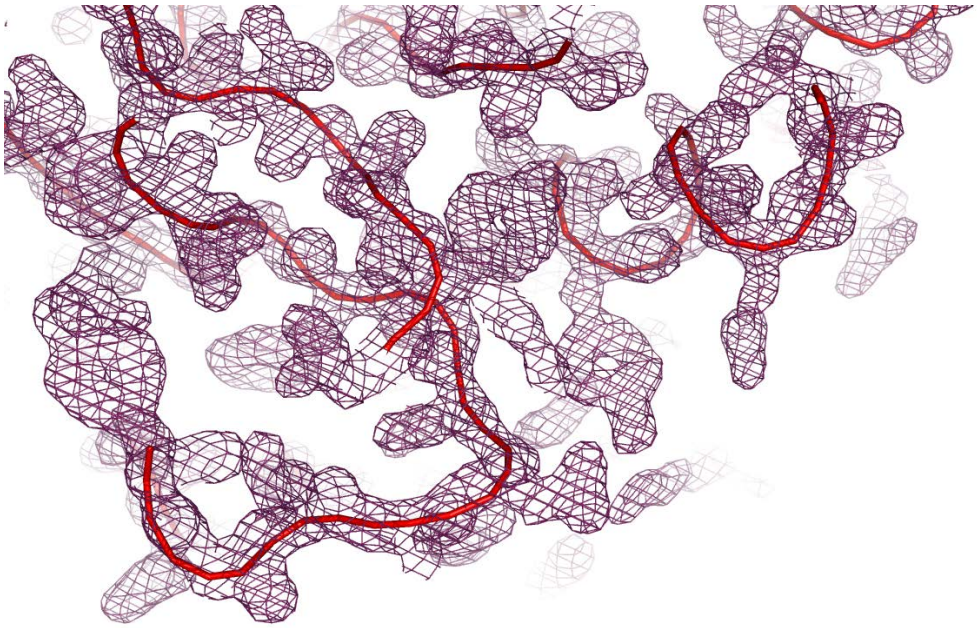


Figure 7.7 – Carte de densité électronique **expérimentale**, améliorée par aplatissement de solvant et contourée à 2,0 σ, obtenue à partir d'un cristal dérivé de CaMalDH avec le GdHPDO3A. Phasage SAD au seuil d'absorption CuKα. 1,7 Å de résolution. Modèle ruban de la CaMalDH superposé.

Contrairement aux données de diffraction des cristaux de PtMalDH et HvMalDH qui ont été enregistrées avec le rayonnement synchrotron, celles, relatives aux cristaux dérivés de CaMalDH obtenus par trempage avec les complexes GdHPDO3A et GdDO3A, ont été enregistrées avec le générateur de

rayons X du laboratoire. Ces données, enregistrées au seuil d'absorption Kα du cuivre, ont conduit à un phasage expérimental permettant la construction automatique de plus de 85 % du modèle avec le complexe GdDO3A et de plus de 95 % avec le complexe GdHPDO3A **[Tableau 7.4 et Figure 7.7]**. Seulement, les cristaux de CaMalDH croissent en présence de 40 mM de chlorure de cadmium. En regardant les cartes de Patterson anomale et considérant le peu de sites Gd construits sur les modèles affinés ainsi que leur taux d'occupation, nous nous sommes demandé si les deux phasages expérimentaux n'avaient pas pu être réalisés du fait de la contribution anomale des ions Cd^{2+}. En effet, à la longueur d'onde correspondant au seuil d'absorption Kα du cuivre (1,541 Å), la valeur de la partie imaginaire *f"* du facteur de diffusion atomique de l'atome de Cd est de 4,71 e^-. De plus, 3 des 10 atomes de Cd construits ont un taux d'occupation supérieur à 60 % (dont un à 100 %). Nous avons donc réalisé une expérience "contrôle" en enregistrant des données de diffraction au seuil d'absorption L_{III} du gadolinium (le générateur de rayon X du laboratoire ayant rendu l'âme) sur un cristal natif de CaMalDH (sans complexe de gadolinium). A cette longueur d'onde (1,711 Å), la partie imaginaire *f"* du facteur de diffusion atomique de l'atome de Cd est de 5,76 e^-, on le "favorise" donc. Le résultat de cette expérience fut non-équivoque : aucun phasage *de novo* n'a pu aboutir alors que les statistiques d'intégration des données (non-présentées ici) montraient aussi que le cristal avait parfaitement supporté l'exposition à la lumière synchrotron. Le phasage expérimental de la CaMalDH est donc bien du à la diffusion anomale des atomes de gadolinium.

Phasage expérimental des données PtMalDH avec le complexe GdHPDO3A

	GdHPDO3A
	PtMalDH
Type de données	$SAD_{Gd_{L_{III}}}$
Résolution (Å)	46.70 - 2.54 (2,68 - 2,54)
Pics$_{Fa}$ **(Pic**$_{Soufre}$**) pk (σ)**	54,88 / 14,05 / 12,26 (9,69)
FOM$^{acen.}_{sharp}$ **/ FOM**$^{cen.}_{sharp}$	0,30 / 0,03
Nombre de sites	3
Résolution FOM$^{acen.}_{sharp} \approx$ **0,5 (Å)**	4,29
P$\varphi^{\neq anomales}_{sharp\,acen.}$ **/ P**$\varphi^{\neq isomorphes}_{sharp\,acen.}$	1,24 / 0,00
SIGMAA$_{solomon}$	0,76
Buccaneer (construit / total)	634 / 648

Tableau 7.5 – Statistiques de phasage expérimental de PtMalDH avec le complexe GdHPDO3A

Phasage par remplacement moléculaire des données HvMalDH avec le complexe GdHPDO3A

	HvMalDH
LLG	5577,4
RFZ	21,2
TFZ	64,8

Tableau 7.6 – Statistiques de phasage par remplacement moléculaire de HvMalDH avec le complexe GdHPDO3A

7.1.3.4 Affinement des structures de CaMalDH, PtMalDH et HvMalDH

Les trois structures obtenues ont été affinées. Elles comportent la modélisation des atomes de lanthanide et, dans la mesure du possible, celle des ligands.

Affinement de la structure de CaMalDH avec modélisation du GdHPDO3A

Huit atomes de lanthanide ont pu être modélisés. Comme nous l'avons indiqué dans l'introduction, il est fréquent que deux molécules de complexe GdHPDO3A s'associent à la surface de la protéine. C'est effectivement ce que la structure révèle : il y a deux sites de fixation par monomère et, dans chacun des sites, le complexe GdHPDO3A se lie par deux.

	GdHPDO3A
	CaMalDH
Résolution (Å)	19,63 - 1,67
R_{work} (%)	15,60
R_{free} (%)	19,36
Nbre de réflexions utilisées	151239
Composition (nbre d'atomes)	
Protéine	4609
H2O	909
Ions	9
Ligands	12
Res. hors Ramachandran (%)	0,33
Ecarts types (globaux)	
Liaisons$_{Longueurs}$ (Å)	0,021
Liaisons$_{Angles}$ (°)	1,637
B_{moyen} protéine (Å^2)	18,38
B_{min} protéine (Å^2)	6,74
B_{max} protéine (Å^2)	110,57
B_{moyen} ions divers (Å^2)	26,21
B_{moyen} ligands divers (Å^2)	33,35
B_{moyen} eau (Å^2)	33,50
B_{moyen} ions lanthanide (Å^2)	35,21
B_{moyen} ligands (Å^2)	34,04
Site 1_{Occ} (%)	53
Site 2_{Occ} (%)	51
Site 3_{Occ} (%)	32
Site 4_{Occ} (%)	29
Site 5_{Occ} (%)	26
Site 6_{Occ} (%)	24
Site 7_{Occ} (%)	14
Site 8_{Occ} (%)	11

Tableau 7.7 – Résumé des statistiques d'affinement pour la structure de CaMalDH avec le complexe GdHPDO3A

Affinement de la structure de PtMalDH avec modélisation du GdHPDO3A

La structure de PtMalDH présente trois sites de fixation pour le complexe GdHPDO3A. Comme cela avait été le cas avec CaMalDH, la construction du modèle confirme non seulement le nombre de sites révélé par la synthèse de Fourier anomale mais aussi leurs taux d'occupation respectifs. En effet, que ce soit pour CaMalDH ou PtMalDH, on remarque que les écarts entre les hauteurs des pics calculées par la synthèse de Fourier anomale sont proportionnelles aux écarts entre les taux d'occupation affinés des sites **[Tableau 7.7 avec tableau 7.4, tableau 7.8 avec tableau 7.5]**.

	GdHPDO3A
	PtMalDH
Résolution (Å)	44,48 - 2,54
R_{work} (%)	18,78
R_{free} (%)	22,39
Nbre de réflexions utilisées	55806
Composition (nbre d'atomes)	
Protéine	4959
H2O	101
Ions	0
Ligands	102
Res. hors Ramachandran (%)	0,93
Ecarts types (globaux)	
Liaisons$_{Longueurs}$ (Å)	0,004
Liaisons$_{Angles}$ (°)	0,726
B_{moyen} protéine (Å^2)	45,64
B_{min} protéine (Å^2)	22,62
B_{max} protéine (Å^2)	144,56
B_{moyen} ligands divers (Å^2)	**55,50**
B_{moyen} eau (Å^2)	43,51
B_{moyen} ions lanthanide (Å^2)	49,33
B_{moyen} ligands (Å^2)	49,20
Site 1$_{Occ}$ (%)	53
Site 2$_{Occ}$ (%)	24
Site 3$_{Occ}$ (%)	23

Tableau 7.8 – Résumé des statistiques d'affinement pour la structure de PtMalDH avec le complexe GdHPDO3A

Affinement de la structure de HvMalDH avec modélisation du GdHPDO3A

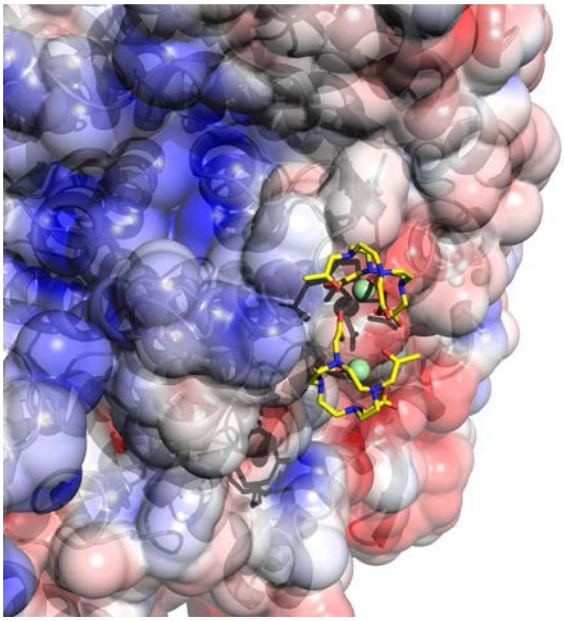
	GdHPDO3A
	HvMalDH
Résolution (Å)	48,40 - 1,95
R_{work} (%)	18,14
R_{free} (%)	21,57
Nbre de réflexions utilisées	179201
Composition (nbre d'atomes)	
Protéine	9160
H2O	787
Ions	28
Ligands	0
Res. hors Ramachandran (%)	1,49
Ecarts types (globaux)	
Liaisons$_{Longueurs}$ (Å)	0,007
Liaisons$_{Angles}$ (°)	0,973
B_{moyen} protéine (Å^2)	37,67
B_{min} protéine (Å^2)	14,24
B_{max} protéine (Å^2)	133,19
B_{moyen} ions divers (Å^2)	56,44
B_{moyen} eau (Å^2)	42,94

Tableau 7.9 – Résumé des statistiques d'affinement pour la structure de HvMalDH avec le complexe GdHPDO3A

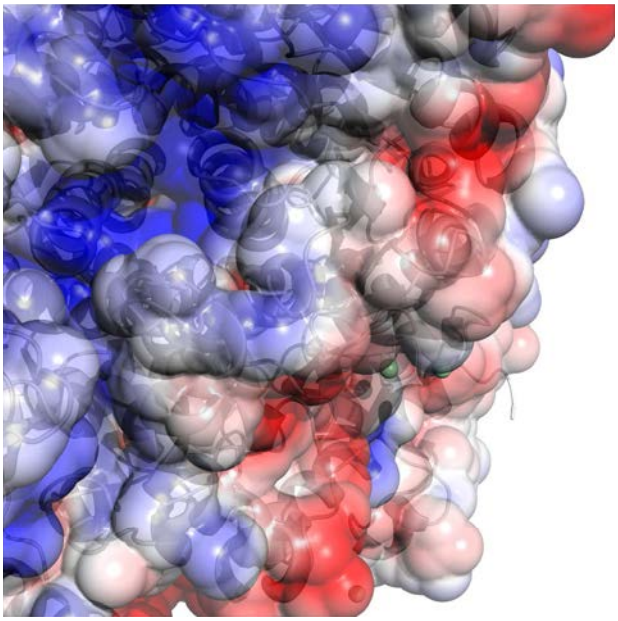
7.1.3.5 Comparaison des surfaces électrostatiques, à région équivalente, des trois MalDH

Dans les travaux décrits dans la première partie, nous avons vu que la composition en acides aminés du site de fixation et la répartition des charges sont des paramètres essentiels pour expliquer le mode de fixation des complexes de lanthanide. Pour illustrer la nouvelle approche apportée par les MalDH, j'ai choisi les premières observations issues de l'étude comparative des surfaces électrostatiques de la région de fixation du complexe GdHPDO3A des trois MalDH .

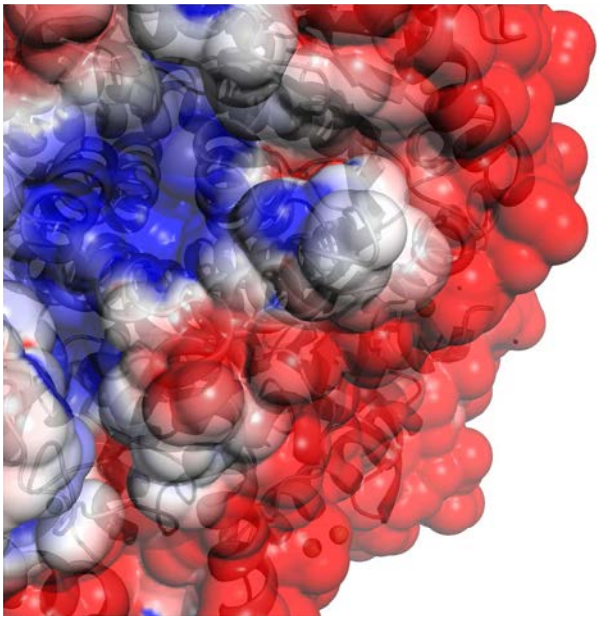
En comparant les surface électrostatiques de la même région sur les trois MalDH différentes, région occupée à 53 % par le complexe GdHPDO3A sur la CaMalDH, on peut déjà observer des différences qui pourraient expliquer la fixation ou l'absence de fixation du complexe **[Figure** 7.8**]**.



(a) Surface électrostatique de la région du 1^{er} site de fixation du GdHPDO3A sur CaMalDH



(b) Surface électrostatique de la région correspondante chez PtMalDH



(c) Surface électrostatique de la région correspondante chez HvMalDH

Figure 7.8 – Surfaces électrostatiques de la région du premier site de fixation du complexe GdHPDO3A sur CaMalDH et de la région correspondante chez PtMalDH et HvMalDH. Echelle de -3.0 $kT.e^{-1}$ (charge négative, rouge) à +3.0 $kT.e^{-1}$ (charge positive, bleu)

On voit que, dans cette région, la charge de la surface de CaMalDH diffère nettement de celle de HvMalDH. Bien plus contrastée et chargée négativement chez HvMalDH, la surface de CaMalDH est, dans cette région, plus hydrophobe. Cela coïncide avec le mode de fixation du complexe GdHPDO3A qui implique des interactions hydrophobes avec les surfaces des protéines (*cf.* introduction).

Le complexe ne s'est pas fixé dans cette région à la surface de PtMalDH. Pourtant, la répartition des charges semble équivalente entre les deux surfaces. Toutefois, on peut émettre l'hypothèse de l'influence de la topologie de la surface qui est très différente entre les deux MalDH. En effet, dans cette région, la surface accessible au solvant de PtMalDH présente une proéminence absente chez CaMalDH.

Au regard des structures secondaires qui composent cette région de la surface de CaMalDH, nous avons constaté que le complexe GdHPDO3A se fixe sur une boucle hydrophobe. Cette boucle existe aussi chez PtMalDH mais, dans le cas présent, sa conformation est différente. L'analyse plus approfondie de ce site de fixation du GdHPDO3A sur CaMalDH révèle la présence d'un ion cadmium qui pourrait expliquer le changement de conformation de la boucle et la fixation du complexe. En effet, à la surface de CaMalDH, cet ion Cd^{2+} crée un pont salin entre la fonction carboxylate d'une des deux conformations de l'acide aminé Glu36 modélisées et l'une des trois fonctions carboxylates du ligand HPDO3A. Les taux d'occupation affinés respectifs la conformation de Glu36 impliquée (45 %), de l'ion Cd^{2+} (45 %) et du complexe GdHPDO3A (51 %) sont en accord avec cette observation. La présence de l'ion Cd^{2+} et du complexe a modifié la conformation générale de cette boucle hydrophobe de CaMalDH, permettant la fixation du complexe. Le fait qu'aucune fixation du complexe GdHPDO3A n'ait pu être constatée chez PtMalDH dans cette région, qui présente les mêmes caractéristiques structurales, illustre bien l'influence des conditions de cristallisation sur la fixation des complexes.

7.1.4 Résumé des premières observations issues de l'étude de l'interaction des trois complexes IRM sur les trois MalDH

Bien qu'ayant seulement choisi l'exemple de l'étude du mode de fixation du GdHPDO3A sur trois MalDH pour illustrer "l'approche MalDH", je consigne ici l'ensemble des premières observations concernant les fixations de trois complexes IRM sur les trois MalDH précitées.

De manière générale, il ne m'a été possible de modéliser avec fiabilité le ligand de ces complexes que pour un taux d'occupation supérieur ou égal à 54 % et pour un facteur d'agitation thermique maximum de 45,00 +/- 5.00 $Å^2$.

Nom de l'organisme dont est issu la MalDH	**Caractère extrêmophile de l'organisme**	**Conditions de cristallisation**	**Gd HPDO3A**	**Gd DO3A**	**Gd DTPABMA**
Chloroflexus aurantiacus	Bactérie thermophile	4 - 14 % PEG 400 100 mM acétate de sodium pH 4,6 40 mM acétate de cadmium	++ (53 %, 51 %)	++	+
Picrophilus torridus	archée hyperthermoacidophile	1,2 - 1,8 M sulfate d'ammonium 100 mM acide citrique pH 3,5/4,0	++ (53 %)	-	+
Haloferax volcanii	archée halophile	34 - 36 % PEG 400 100 mM Tris HCl pH 8,0 100 mM chlorure de magnésium	+	++ (65 %, 56 %, 44 %)	++ (83 %, 59 %, 59 %, 50 %, 48 %, 46 %, 41 %)

Tableau 7.10 – Résumé des résultats de l'étude de l'interaction de trois complexes IRM sur les MalDH. (**-**) Aucune fixation, (+) fixation faible, site détecté par synthèse de Fourier anomale, (++) fixation forte, phasage *de novo* par méthode SAD possible. Seuls les taux d'occupation affinés des sites occupés à plus de 40 % sont indiqués

Le **tableaux 7.10** montre que la fixation des complexes GdHPDO3A, GdDO3A et GdDTPA-BMA diffère selon le complexe et la MalDH utilisés.

Le mode de fixation des complexes de lanthanide, utilisés jusqu'alors, a révélé l'importance des liaisons hydrogènes et des liaisons ioniques qui se créent entre les acides aminés de la surface protéique et les complexes. Les complexes utilisent les groupements fonctionnels ionisables de ces acides aminés pour créer ce type de liaison : liaisons hydrogènes entre la fonction acide des ligands dipicolinates et la fonction amine des acides aminés arginine et lysine, par exemple, ou liaisons ioniques entre l'ion lanthanide et le groupement acide des acides aminés aspartate et glutamate dans le cas du GdDO3A.

Mais tous ces groupements fonctionnels ionisables ont un pKa en solution de : 3,90 pour la fonction acide de la chaine latérale des aspartates, 4,30 pour la fonction acide de la chaine latérale des glutamates, 10,54 pour la fonction amine de la chaine latérale des lysines et 12,48 pour la fonction amine tertiaire de la chaine latérale des arginines. Bien qu'il puisse y avoir une distinction entre le pKa de ces fonctions et le pKa local (mesuré sur un groupe d'acides aminés couvrant une zone donnée), la protonation de ces acides aminés est toujours dépendante du pH de la solution tampon utilisée dans les conditions de cristallisation. De fait, la fixation des complexes doit être dépendante de cette protonation.

L'influence du pH pourrait par exemple expliquer l'absence totale de fixation du complexe GdDO3A sur l'enzyme PtMalDH. Les essais d'obtention de cristaux dérivés par co-cristallisation ont été réalisés avec une solution tampon d'acide citrique à pH 3,5. D'une part, à ce pH, tous les groupements acides des aspartates et des glutamates sont protonés. D'autre part, le complexe LnDO3A est le seul complexe qui se fixe aux surfaces protéiques, non par son ligand mais par l'ion lanthanide. Ainsi, de part leur état carboxylique et non plus carboxylate, il est possible que ces fonctions ne puissent plus être impliquées dans une liaison ionique avec l'ion lanthanide du complexe LnDO3A.

CaMalDH -MRKKISIIGAG-FVGSTTAHWLAAKELGDIVLLDFVEGVPQGKALDLYEASPIEGFDVR 58
PtMalDH MARSKISVIGAG-AVGATVAQTLAIRQTGDIYIFDIVDGLAEGKALDILEGAPHWGYDLD 59
HvMalDH --MTKVSVIGAAGTVGAAGGYNLALRDVCDELVFVDIPKMEDKTVGQAADTNHGIAYDSN 58
CaMalDH VTG-----TNNYADTANSDVIVVTSGAPRKPGMSREDLIKVNADITRACISQAAPLSPNA 113
PtMalDH IKGFCTADESKYAEMKGSDVIVVTAGLARKPGMSRDDLLLKNIGIMKSVGEAIKKYSPES 119
HvMalDH TVVT----QGGYEDTAGSDVVVITAGIPRQPGQTRIDLAGDNAPIMDDIGSSLAEYNDDF 114
CaMalDH VIIMVNNPLDAMTYLAAEVSGFPKERVIGQAGVLDAARYRTFIAMEAGVSVEDVQAMLMG 173
PtMalDH KIVVVTNPADIMAYAIYKASGISPERIIGLGGSLDSTRFRTFLAQELNVSFEDVNAFVIG 179
HvMalDH VSITTSNPVDLLNRHLYETGDRDRDRHKVIGFGGRLDSARFRYVLSQRFDVPVKNVDATILG 174
CaMalDH GHGDEMVPLPRFSTISGIPVSEFIAPDRLAQIVERTRKGGGEIVNLLKTGSAYYAPAAAT 233
PtMalDH GHGDDMVPFIRYSNVSGIPIEDLLPREKIDEIVKRTRFGGGEIVNLYKTGSAFYAPGISI 239
HvMalDH EHGDAQVPVFSKVRVDGN--DPAFSADEKEEILGDLQESAMDVIE--RKGATQWGPATGV 230
CaMalDH AQMVEAVLKDKKRVMPVAAYLTGQYG----LNDIYFGVPVILGAGGVEKILELPLNEEEM 289
PtMalDH AVMVESIVNDRKRVIPCAAYITGEHSKTYLVNNLFIGVPIKIGKNGVEKIYELKFNEDEL 299
HvMalDH AHMVEAVLHDTGEVLPGSLVLDGEFG----YEDTAFGVPVKLGSNGIEEVVEWDLDDYAD 286
CaMalDH ALLNASAKAVR---ATLDTLKSL-- 309
PtMalDH EAWKKSVESVKKNSAIADDYFAKNQ 324
HvMalDH LMNDAAEKLLDQYDEIA-------- 303
KEDR Site de fixation du complexe GdHPDO3A
KEDR Site de fixation du complexe GdDO3A
KEDR Site de fixation du complexe GdDTPABMA
* Acide aminé conservé
: Substitution au sein du même groupe d'acides aminés (polaires ou apolaires) mais conservation des propriétés physico-chimiques
. Substitution au sein du même groupe d'acides aminés sans conservation des propriétés physico-chimiques

Figure 7.9 – Position des sites de fixation des complexes IRM occupés à plus de 40 % ramenés aux séquences primaires d'un monomère de MalDH alignées avec le programme Clustal W

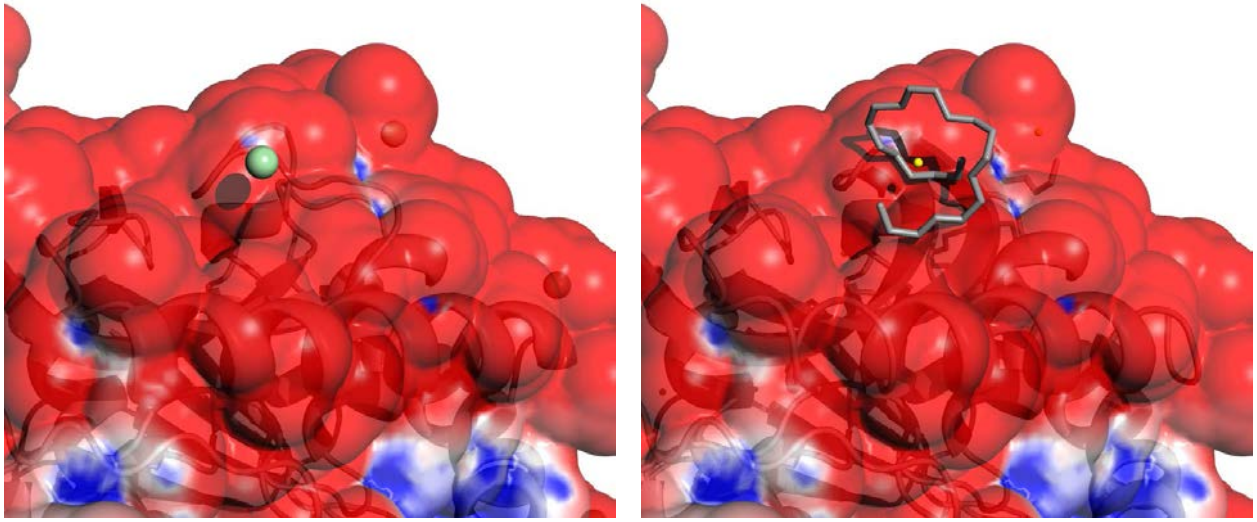
Comme le représente la **[Figure 7.9]**, en s'intéressant tout d'abord aux sites de fixation les plus fortement occupés, on constate que des complexes partagent des sites communs de fixation à la surface d'une même MalDH (cas du GdDO3A et du GdDTPA-BMA à la surface de HvMalDH).

Aligner les séquences primaires permet de mettre aussi en évidence les différences au niveau de la composition en acides aminés des régions impliquées dans la fixation des complexes. Prenons, une fois de plus, l'exemple du GdHPDO3A. On remarque que le GdHPDO3A se fixe sur PtMalDH dans une région équivalente à celle où se fixent, sur HvMalDH, les complexes GdDO3A et GdDTPA-BMA, mais il ne se fixe pas pour autant dans cette région sur HvMalDH. L'inverse est aussi observable : il y a une région de la séquence primaire qui correspond à un site de fixation du GdHPDO3A sur CaMalDH mais qui montre que sur PtMalDH et HvMalDH, il ne s'est pas fixé **[Figure 7.9]**. Ainsi, comparer les compositions en acides aminés respectives des régions dans lesquelles un complexe s'est fixé pourra apporter de nombreuses informations quant au mode de fixation de ces complexes.

En comparant la structure de la protéine HvMalDH native avec celle issue d'un cristal dérivé avec le GdDO3A, on met en évidence que dans un cristal natif, une molécule de PEG occupe (à 100 %) le même site qu'un complexe GdDO3A dans un cristal dérivé. Or, les cristaux dérivés de HvMalDH avec

le GdDO3A ayant été obtenus par trempage, il n'est pas exclu que ce complexe entre en compétition avec la molécule de PEG pour se fixer à ce site. Ceci pourrait d'ailleurs expliquer le taux d'occupation plus faible du complexe à ce site (32 %) **[Figure 7.10]**.

Aussi, nous avons constaté que l'accessibilité de certains sites de fixation pouvait être modifiée par l'empilement cristallin dans le cadre d'une expérience de trempage. HvMalDH étant homotétramérique, le complexe GdDO3A devrait trouver le même site de fixation principal sur chaque monomère de la protéine, soit quatre fois le même site de fixation. Or, de par l'empilement cristallin, le complexe ne se fixe en trempage qu'à trois de ces quatre sites principaux. Le quatrième site est masqué car il est pris dans un contact cristallin.



(a) Surface électrostatique du troisième site de fixation du GdDO3A sur HvMalDH. Echelle de -3.0 kT.e^{-1} (charge négative, rouge) à +3.0 kT.e^{-1} (charge positive, bleu)

(b) Surface électrostatique de même région dans un cristal natif de HvMalDH. Echelle de -3.0 kT.e^{-1} (charge négative, rouge) à +3.0 kT.e^{-1} (charge positive, bleu)

Figure 7.10 – Région du troisième site de fixation du GdDO3A dans un cristal dérivé de HvMalDH (a) et dans un cristal natif de HvMalDH (b)

Si cette approche devrait nous mener à affiner nos connaissances sur les modes de fixation des complexes IRM, elle a déjà permis de confirmer certains modes de fixation déjà déterminés (*cf.* introduction). Par exemple, résoudre la structure de l'enzyme HvMalDH avec le GdDO3A à une résolution de 1.85 Å permet de confirmer que ce complexe se fixe principalement à la protéine via l'atome de lanthanide **[Figure 7.11]** et non via le ligand chimique, comme c'est le cas avec les complexes GdHPDO3A et GdDTPA-BMA (*cf.* introduction). En effet, c'est bien la fonction acide de la chaine latérale des acides aminés aspartates ou glutamates de la protéine qui permet la fixation du complexe entier, en venant compléter la sphère de coordination du lanthanide.

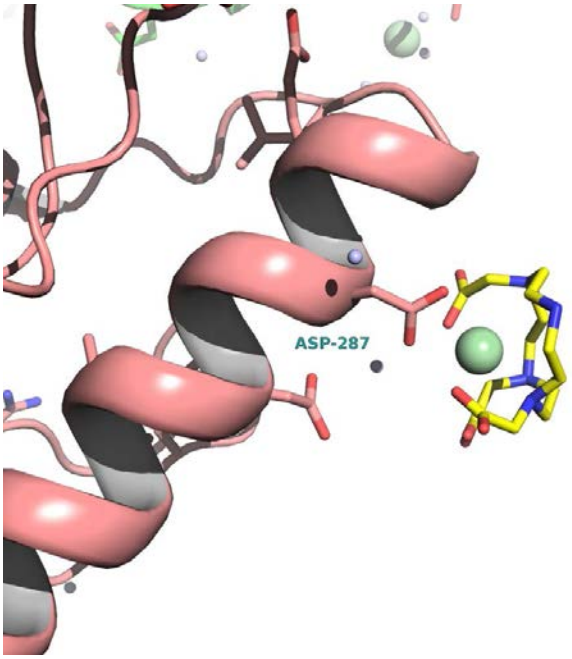


Figure 7.11 – Mode de fixation du complexe GdDO3A à la surface de HvMalDH, site occupé à 72 %

De même, résoudre la structure de CaMalDH avec le GdHPDO3A à la résolution de 1,67 Å nous confirme que ce complexe se fixe toujours par paire **[Figure 7.8]**, comme c'était déjà le cas sur le lysozyme de blanc d'œuf de poule (*cf.* introduction).

Utilisés en co-cristallisation avec les enzymes CaMalDH et HvMalDH, les complexes GdHPDO3A, GDDO3A et GdDTPA-BMA ont empêché la cristallisation. Pourquoi dans certains cas, la présence des complexes de lanthanide dans la goutte empêche-t-elle la cristallisation ? Les complexes de lanthanide utilisés en co-cristallisation peuvent-ils aussi avoir un effet solubilisant ? Se fixent-ils précisément dans les zones de la surface protéique normalement impliquées dans les contacts cristallins, empêchant ainsi la formation de cristaux dérivés et natifs ? Nous n'avons pas encore répondu à ces questions. Fort heureusement, les cas où les complexes de lanthanide ont un "effet solubilisant" sont rares. Et comme nous avons pu le vérifier, lorsque cela se produit, il est toujours possible d'avoir recours au trempage pour obtenir des cristaux dérivés.

7.2 Les nouvelles structures de MalDH résolues par phasage *de novo*

Dans les sections suivantes, je vais brièvement présenter ce que les trois nouvelles structures, résolues par la méthode de phasage SIRAS, ont apporté à la compréhension de l'adaptation halophile.

Comme cela a été expliqué dans l'introduction de ce chapitre, les enzymes MalDH doivent faire face à des conditions de stress dans le cytoplasme de l'organisme. Conditions qui dépendent du caractère extrêmophile de l'organisme. En ce qui concerne les MalDH des cytoplasmes d'organismes halophiles, ces enzymes doivent rester actives à des concentrations en sel supérieures au molaire (Ginzburg and Sachs, 1970). L'étude de la composition et de la structure du solvant à la surface des MalDH halophiles devrait apporter de précieuses informations sur l'adaptation des protéines aux hautes concentrations en sel. En effet, l'hypothèse de base à toute cette étude structurale est que les substitutions adaptatives de

surface chez les protéines halophiles induisent une organisation du solvant à leur surface, différente de celles non-halophiles. Ceci leur confére des propriétés de solubilité très différentes (ce postulat émane d'une étude dont la description détaillée se trouve dans notre article, ajouté à la fin du ce chapitre).

7.2.1 HvMalDH, la malate déshydrogénase de *Haloferax volcanii*

La malate déshydrogénase homotétramérique halophile de *Haloferax volcanii* est composée de 1216 acides aminés (4x304). Sa masse molaire est de 130 kDa et l'analyse de sa composition permet de calculer un point isoélectrique théorique de 4.20 et un coefficient d'extinction molaire théorique ε de 24410 $M^{-1}.cm^{-1}$.

La HvMalDH est purifiée et cristallisée en présence d'ions différents : en présence d'ions potassium à très forte concentration lors de la purification et en présence d'ions magnésium lors de la cristallisation. Dans le cadre de l'étude de l'adaptation halophile, sa structure constitue un bon modèle pour étudier la conservation phylostructurale des potentiels sites ioniques et analyser l'influence de ces ions sur la structure de l'enzyme.

7.2.1.1 Intégration des données en vue d'un phasage par la méthode SIRAS

Un jeu de données de diffraction a été enregistré au seuil d'absorption L_{III} du gadolinium à partir d'un cristal dérivé de HvMalDH obtenu par trempage dans une solution de complexe GdDTPA-BMA. L'autre jeu de données est issu d'un cristal natif et a été enregistré au seuil d'absorption K du sélénium à une résolution de 1,5 Å.

	HvMalDH	
Source synchrotron	ESRF	
Ligne de lumière	ID23-1	FIP-BM30A
Méthode	**SIRAS**	
Données	Jeu anomal	jeu natif
λ (Å)	1,711 ($Gd_{L_{III}}$)	0,979 ($Se_{L_{III}}$)
Groupe d'espace	C2	
Paramètres de maille (Å,°)	a = 143,750, b = 82,320, c = 112,750	a = 143,680, b = 83,060, c = 111,940
	ß = 101,720	ß = 102,020
Contenu unité asymétrique	tétramère	
Résolution (Å)	41,58 - 1,95 (2,05 - 1,95)	48,34 - 1,50 (1,58 - 1,50)
Nombre de réflexions uniques	92636 (13029)	205562 (29939)
R_{merge} (%)	11,1 (42,4)	5,8 (40,4)
R_{pim} (%)	7,2 (24,1)	2,3 (16,0)
Sig_{ano} XDS	1,29 (0,85)	-
R_{ano} (%)	9,7 (23,8)	-
$\frac{I}{\sigma(I)}$	8,9 (3,5)	21,6 (4,9)
Complétude (%)	98,6 (95,4)	100,0 (99,9)
Redondance	5,0 (4,9)	7,4 (7,3)

Tableau 7.11 – Statistiques d'intégration des jeux de données utilisés pour le phasage de HvMalDH par la méthode SIRAS ; colonne de gauche : cristal dérivé de HvMalDH avec le complexe GdDTPA-BMA ; colonne de droite : cristal natif de HvMalDH

7.2.1.2 Phasage par la méthode SIRAS

Le phasage, s'appuyant d'une part sur le signal anomal, apporté par le jeu issu du cristal dérivé avec le complexe GdDTPA-BMA et d'autre part sur les différences isomorphes a conduit à la construction automatique de plus de 98 % des acides aminés qui composent le tétramère contenu dans l'unité asymétrique.

	HvMalDH
	SIRAS
Résolution (Å)	50,90 - 1,50
FOM$^{acen.}_{sharp}$ / **FOM**$^{cen.}_{sharp}$	0,15 / 0,06
Nombre de sites	11
Résolution FOM$^{acen.}_{sharp} \approx$ **0,5 (Å)**	4,8
P$\varphi^{\neq isomorphes}_{sharp\,acen.}$ / **P**$\varphi^{\neq anomales}_{sharp\,acen.}$	0,47 / 1,45
Solvant (%)	40
SIGMAA$_{solomon}$	0,69
Buccaneer (construit / total)	1193 / 1216

Tableau 7.12 – Statistiques du phasage SIRAS de HvMalDH avec le complexe GdDTPA-BMA

7.2.1.3 Affinement de la structure de HvMalDH

L'affinement a été effectué avec le programme Phenix selon le protocole indiqué dans l'introduction de ce manuscrit. Les statistiques d'affinement sont consignées dans le tableau suivant.

	HvMalDH
Code RCSB-PDB	4BGU
Résolution (Å)	48,48 - 1,49
R$_{work}$ **(%)**	15,58
R$_{free}$ **(%)**	17,88
Nbre de réflexion utilisées	205541
Composition (nbre d'atomes)	
Protéine	9360
H2O	1814
Ions	94
Ligands	167
Res. hors Ramachandran (%)	0,50
Ecarts types (globaux)	
Liaisons$_{Longueurs}$ **(Å)**	0,010
Liaisons$_{Angles}$ **(°)**	1,300
B$_{moyen}$ **protéine (Å**^{2}**)**	14,24
B$_{min}$ **protéine (Å**^{2}**)**	4,81
B$_{max}$ **protéine (Å**^{2}**)**	63,52
B$_{moyen}$ **eau (Å**^{2}**)**	29,60
B$_{moyen}$ **ions divers (Å**^{2}**)**	22,32
B$_{moyen}$ **ligands divers (Å**^{2}**)**	30,34

Tableau 7.13 – Résumé des statistiques d'affinement de la structure native de HvMalDH

A la fin de l'affinement du modèle, un certain nombre d'ions Mg^{2+} et Cl^{-} ont pu être modélisés. Mais, même si on se fie à la forme anisotrope de la densité électronique et à l'environnement protéique, il est parfois difficile de distinguer un ion Mg^{2+} (14 électrons) ou un ion Cl^{-} (16 électrons) d'une molécule d'eau (10 électrons) si leur taux d'occupation n'est pas très élevé.

7.2.1.4 Une modélisation fiable des ions dans la structure de HvMalDH grâce à la diffusion anomale

Si la densité électronique seule ne peut parfois pas suffire à distinguer des ions d'une molécule d'eau, leur contribution anomale peut permettre de les discriminer **[Tableau 7.14]**.

	$Br_{seuil\ K}$ (λ = 0,921)	**$Gd_{seuil\ L_{III}}$ (λ = 1,711)**
f"$_{Br}$ (e^-)	6,7	-
f"$_{Cl}$ (e^-)	-	0,9
f"$_{Mg}$ (e^-)	0,1	0,2
f"$_{K}$ (e^-)	0,4	1,3
f"$_{S}$ (e^-)	0,2	0,7
f"$_{Mn}$ (e^-)	1,2	3,3
f"$_{Fe}$ (e^-)	1,3	3,8
f"$_{Zn}$ (e^-)	2,0	0,8
f"$_{Ca}$ (e^-)	0,5	1,5

Tableau 7.14 – Valeur de la partie imaginaire du facteur de diffusion atomique des atomes pouvant se trouver dans un cristal de HvMalDH, en fonction de la longueur d'onde utilisée.

C'est dans cette optique que nous avons mené une nouvelle expérience de diffraction avec un cristal de HvMalDH ayant été obtenu, non plus en présence de $MgCl_2$ mais en présence de $MgBr_2$. Alors qu'il n'est pas possible d'atteindre le seuil d'absorption K du chlore sur une ligne de lumière synchrotron, il est possible de travailler à une longueur d'onde qui correspond au seuil d'absorption K du brome. Nous avons supposé que les ions Br^{-} se comportent comme les ions Cl^{-} à la surface de la protéine, aux vues de leur propriétés (Wells, 1984), **[Tableau 7.15]**.

	HvMalDH	
	Cristaux en $MgBr_2$	**Cristaux en $MgCl_2$**
λ (Å)	0,921 (Br_K)	1,711 ($Gd_{L_{III}}$)
Groupe d'espace	C2	
Paramètres de maille (Å,°)	a = 144,886, b = 83,475, c = 112,377 ß = 101,535	a = 144,610, b = 83,418, c = 112,311 ß = 101,505
Contenu unité asymétrique	tétramère	
Résolution (Å)	48,45 - 1,60 (1,69 -1,60)	46,00 - 1,88 (1,98 - 1,88)
Nombre de réflexions uniques	170600 (24312)	99062 (13575)
R_{merge} (%)	5,4 (41,5)	5,0 (10,2)
R_{pim} (%)	4,4 (34,4)	4,3 (8,2)
R_{ano} (%)	3,4 (25,5)	3,9 (6,5)
$\frac{I}{\sigma(I)}$	10,3 (1,9)	8,9 (6,6)
Complétude (%)	98,8 (96,7)	93,0 (87,5)
Redondance	4,6 (4,6)	4,0 (3,9)

Tableau 7.15 – Statistiques d'intégration du jeu de données HvMalDH enregistré au seuil d'absorption K du brome sur un cristal en présence de $MgBr_2$ (colonne de gauche) et du jeu de données enregistré au seuil d'absorption L_{III} du gadolinium sur un cristal natif obtenu en présence de $MgCl_2$ (colonne de droite)

La diffusion anomale des atomes de brome nous permet donc de confirmer la position des atomes de chlore. De plus, un jeu enregistré au seuil d'absorption L_{III} du gadolinium enregistré sur un cristal natif obtenu en présence de $MgCl_2$ nous renseigne sur la diffusion anomale de ces atomes de chlore, même si celle-ci est faible. Enfin, en ce qui concerne les ions Mg^{2+}, les deux enregistrements effectués (au seuil d'absorption K du brome et au seuil d'absorption L_{III} du gadolinium) conduisent à une valeur équivalente (faible) de la contribution imaginaire f'' pour les atomes de magnésium **[Figure 7.12]**.

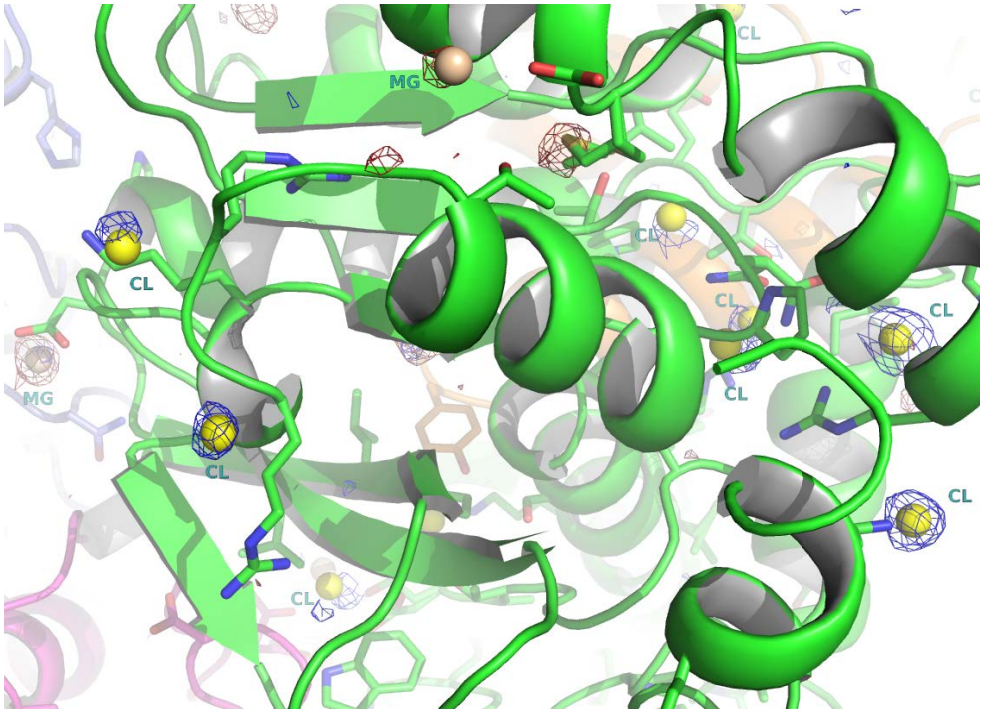


Figure 7.12 – Cartes de densité anomales calculées avec le jeu de données enregistré au seuil d'absorption K du Brome (en bleu) et avec le jeu de données enregistré au seuil d'absorption L_{III} du gadolinium (en rouge) dans la structure de HvMalDH

Des synthèses de Fourier anomale, utilisant le modèle affiné, ont été calculées avec le programme *ANODE* (Thorn and Sheldrick, 2011) et ont permis de modéliser avec assurance 71 ions Cl^- et 23 ions K^+ (position, taux d'occupation et facteur d'agitation thermiques affinés) dans la structure de HvMalDH. En effet, nous nous sommes aperçu, aposteriori, que la différence des hauteurs des pics calculées pour chacune des deux longueur d'ondes correspondait finalement, non pas à celle d'atomes de magnésium mais à celle d'atomes de potassium. Nous nous attendions pourtant a modéliser des ions magnésium car ils sont présents dans les conditions de cristallisation de la protéine. Mais compte tenu du fait du protocole de purification qui requiert l'utilisation de tampons très concentrés en chlorure de potassium, il était tout aussi logique de trouver des ions K^+ dans la protéine HvMalDH.

7.2.2 PtMalDH, la malate déshydrogénase de *Picrophilus torridus* et la localisation de la boucle catalytique

La malate déshydrogénase homotétramérique de *Picrophilus torridus* est composée de 1296 acides aminés (4x324). Cette enzyme hyperthermoacidophile a une masse molaire de 140 kDa et l'analyse de sa composition donne un point isoélectrique théorique de 5,80 et un coefficient d'extinction molaire théorique ε de 30495 $M^{-1}.cm^{-1}$.

Même si ce que je présente à propos de PtMalDH ne rentre pas tout à fait dans le cadre de l'étude de l'adaptation halophile, puisque *Picrophilus torridus* est un organisme adapté au bas pH, cela permet néanmoins de présenter les nouvelles structures et l'information qu'elles apportent quant à la fonction enzymatique des MalDH.

A partir des deux conditions de cristallisation possibles obtenues et décrites dans la sous-section 7.1.1 page 214, deux types de cristaux de groupe d'espace de symétrie quadratique ont pu être obtenus, conduisant à l'obtention de deux modèles différents.

7.2.2.1 Intégration des données de diffraction

Par la méthode SIRAS, nous avons résolu la structure du dimère A-D de PtMalDH. Pour cette méthode, nous avons utilisé un jeu de données anomal enregistré à partir d'un cristal dérivé co-cristallisé avec le complexe GdHPDO3A et un jeu de données enregistré à partir d'un cristal natif. Je précise, pour la suite, que ces cristaux ont été obtenus dans une condition de cristallisation contenant du sulfate d'ammonium **[Tableau 7.16]**.

	PtMalDH	
	Cristaux obtenus avec sulfate d'ammonium	
Source synchrotron	SOLEIL	
Ligne de lumière	PROXIMA1	
Méthode	**SIRAS**	
Données	Jeu anomal	jeu natif
λ (Å)	1,711 ($Gd_{L_{III}}$)	0,979 ($Se_{L_{III}}$)
Groupe d'espace	$I4_122$	
Paramètres de maille (Å)	a = 112,402, c = 286,400	a = 112,211, c = 287,261
Contenu unité asymétrique	dimère A-D	
Résolution (Å)	47,73 - 2,54 (2,68 - 2,54)	49,43 - 2,10 (2,21 -2,10)
Nombre de réflexions uniques	29874 (3601)	53817 (7581)
$\mathbf{R}_{merge}$ **(%)**	5,8 (29,7)	8,2 (55,3)
$\mathbf{R}_{pim}$ **(%)**	3,5 (19,5)	3,1 (20,8)
$\mathbf{Sig}_{ano}$ **XDS**	1,43 (0,75)	-
$\mathbf{R}_{ano}$ **(%)**	4,9 (19,7)	-
$\frac{I}{\sigma(I)}$	11,4 (2,6)	7,6 (1,4)
Complétude (%)	97,3 (81,1)	99,6 (97,6)
Redondance	6,7 (4,8)	8,0 (8,0)

Tableau 7.16 – Statistiques d'intégration des jeux de données utilisés pour le phasage de PtMalDH par la méthode SIRAS ; colonne de gauche : d'un cristal dérivé de PtMalDH avec le complexe GdHPDO3A ; colonne de droite : d'un cristal natif de PtMalDH., cristaux obtenus dans les conditions avec sulfate d'ammonium

Comme nous allons le voir, suite à une demande du Dr. Madern, un autre jeu de données a été enregistré sur un cristal natif ayant été obtenu à partir des conditions de cristallisation sans sulfate d'ammonium (en présence de PEG) **[Tableau 7.17]**.

	PtMalDH
	Cristal obtenu sans sulfate d'ammonium
Source synchrotron	ESRF
Ligne de lumière	FIP-BM30A
Méthode	Remplacement moléculaire
Données	jeu natif
λ (Å)	0,979 (Se_K)
Groupe d'espace	$P4_12_12$
Paramètres de maille (Å)	a = 81,167, c = 395,991
Contenu unité asymétrique	tétramère complet
Résolution (Å)	46,47 - 1,81 (1,91 -1,81)
Nombre de réflexions uniques	121716 (17271)
$\mathbf{R}_{merge}$ **(%)**	11,9 (67,3)
$\mathbf{R}_{pim}$ **(%)**	3,6 (21,1)
$\frac{I}{\sigma(I)}$	6,3 (1,3)
Complétude (%)	99,8 (98,5)
Redondance	11,5 (10,8)

Tableau 7.17 – Statistiques d'intégration du jeu de données de diffraction d'un cristal natif obtenu en présence de PEG

7.2.2.2 Calcul des phases pour les deux types de données enregistrées

Le phasage s'appuyant d'une part sur le signal anomal, apporté par le jeu issu du cristal dérivé avec le complexe GdHPDO3A, et d'autre part sur les différences isomorphes, a conduit à la construction d'un premier modèle **[Tableau 7.18]**. Une fois ce modèle affiné, nous l'avons utilisé pour réaliser un phasage par remplacement moléculaire avec les données issues du cristal natif obtenu sans sulfate d'ammonium (le protocole de remplacement moléculaire qui s'en suivit est indiqué dans l'introduction) **[Tableau 7.19]**.

	PtMalDH
	SIRAS
	Avec $(NH_4)_2SO_4$
Résolution (Å)	49,43 - 2,10
FOM$^{acen.}_{sharp}$ / **FOM**$^{cen.}_{sharp}$	0,14 / 0,09
Nombre de sites	3
Résolution FOM$^{acen.}_{sharp} \approx$ **0,5 (Å)**	6,13
P$\varphi^{\neq isomorphes}_{sharp\,acen.}$ / **P**$\varphi^{\neq anomales}_{sharp\,acen.}$	0,40 / 0,95
Solvant (%)	52
SIGMAA$_{solomon}$	0,64
Buccaneer (construit / total)	508 / 648

Tableau 7.18 – Statistiques du phasage expérimental SIRAS de PtMalDH avec le complexe GdHPDO3A. Cristaux obtenus en présence de sulfate d'ammonium "$(NH_4)_2SO_4$"

	PtMalDH
	Sans $(NH_4)_2SO_4$
LLG	6236
RFZ	42,6
TFZ	106,8

Tableau 7.19 – Statistiques du phasage par remplacement moléculaire. Cristaux obtenus sans sulfate d'ammonium

7.2.2.3 Affinement des structures de PtMalDH

L'affinement a été effectué avec le programme Phenix selon le protocole indiqué dans l'introduction de ce manuscrit. Les statistiques d'affinement sont consignées dans les tableaux suivants.

	PtMalDH
	Avec $(NH_4)_2SO_4$
Résolution (Å)	49,43 - 2,10
R_{work} (%)	18,52
R_{free} (%)	20,90
Nbre de réflexion utilisées	53717
Composition (nbre d'atomes)	
Protéine	4992
H2O	329
Ligands	265
Res. hors Ramachandran (%)	0,60
Ecarts types (globaux)	
Liaisons$_{Longueurs}$ (Å)	0,003
Liaisons$_{Angles}$ (°)	0,713
B_{moyen} protéine (Å^2)	41,15
B_{min} protéine (Å^2)	17,64
B_{max} protéine (Å^2)	148,01
B_{moyen} eau (Å^2)	46,79
B_{moyen} ligands divers (Å^2)	53,81

Tableau 7.20 – Résumé des statistiques d'affinement de la structure native de PtMalDH avec construction d'ions sulfates "SO_4" et NAD^+

A l'issue de l'affinement de la première structure, nous avons été à même de construire un modèle de l'enzyme dans son intégralité. Nous avons aussi pu construire un certain nombre de molécules de sulfate ainsi qu'une molécule de cofacteur NAD^+. Le NAD^+ n'ayant pas été ajouté, ni au cours de la purification ni au cours de la cristallisation, les deux molécules de NAD^+ piégées dans chaque monomère doivent provenir du cytoplasme de la bacterie *E. coli*, utilisée pour la production des MalDH recombinantes.

A la demande du Dr. Madern qui souhait aussi avoir une version dépourvue de tout ligand ("forme apo"), nous avons donc résolu la structure de l'enzyme entière issue du cristal obtenu dans les conditions sans sulfate d'ammonium. En effet, cette fois, aucun groupement sulfate n'a pu être modélisé ni aucune molécule de NAD^+.

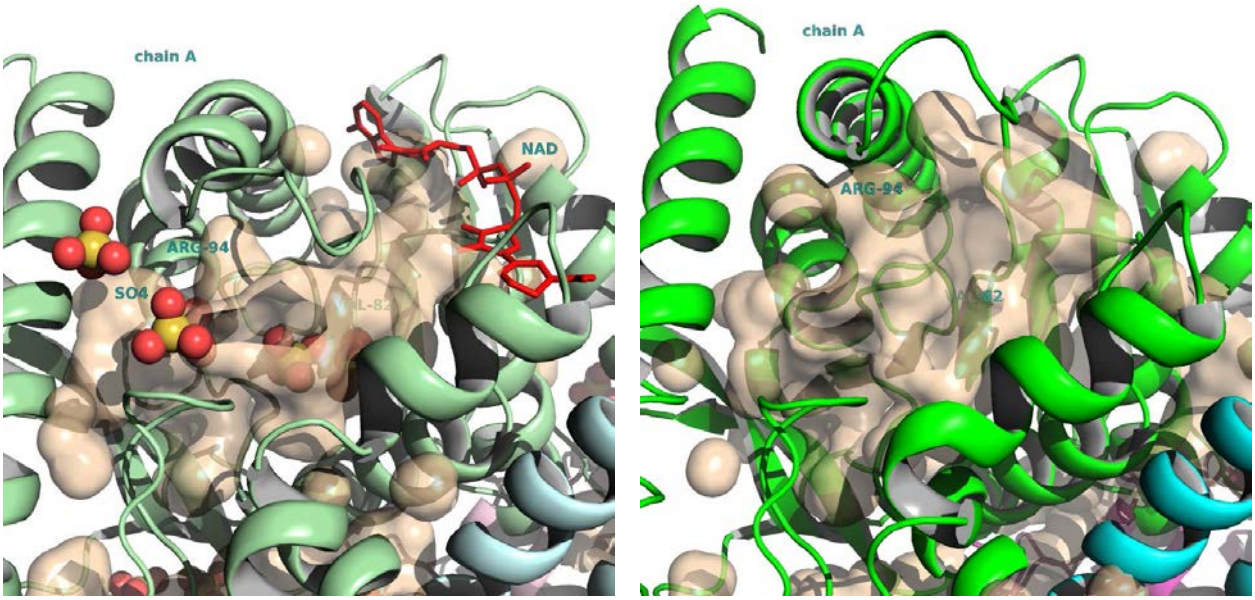
	PtMalDH
	Sans $(NH_4)_2SO_4$
Code RCSB-PDB	4BGV
Résolution (Å)	46,47 - 1,81
R_{work} (%)	14,65
R_{free} (%)	18,47
Nbre de réflexion utilisées	121529
Composition (nbre d'atomes)	
Protéine	10046
H2O	1219
Ligands	100
Res. hors Ramachandran (%)	0,60
Ecarts types (globaux)	
Liaisons$_{Longueurs}$ (Å)	0,008
Liaisons$_{Angles}$ (°)	1,097
B_{moyen} protéine (Å^2)	17,95
B_{min} protéine (Å^2)	5,24
B_{max} protéine (Å^2)	145,18
B_{moyen} eau (Å^2)	31,05
B_{moyen} ligands divers (Å^2)	31,25

Tableau 7.21 – Résumé des statistiques d'affinement de la structure native de PtMalDH sous forme apo

7.2.2.4 Les deux conformations de la boucle catalytique de PtMalDH

Le site catalytique des malate et des lactate déshydrogénases se compose d'une crevasse catalytique recouverte par une boucle mobile. Les acides aminés impliqués dans la fixation du substrat et la réactivité chimique sont très conservés [**Figure 7.9 page 229]**. Le site catalytique des malate et des lactate déshydrogénases se compose d'une crevasse catalytique recouverte par une boucle mobile. Les acides aminés impliqués dans la fixation du substrat et la réactivité chimique sont très conservés. La boucle mobile (située entre les acides aminés 82 et 94) porte l'acide aminé discriminant la reconnaissance du substrat entre malate et lactate déshydrogénases. Dans le cas des MalDH, il s'agit d'une arginine conservée en position 83 qui établit une liaison avec une des deux extrémités carboxyles de l'oxaloacétate et stabilise ainsi la liaison du substrat avec l'arginine 152. Dans le cas des LacDH, on trouve une glutamine en position 83. L'absence d'une charge positive ne permet pas la stabilisation de l'oxaloacétate, en revanche cela autorise malgré tout la fixation du pyruvate par l'arginine 152 (Madern, 2002).

Depuis le début de cette section, si j'ai distingué la structure issue d'un cristal ayant été obtenu en présence de sulfate d'ammonium de celle issue d'un cristal obtenu dans des conditions sans sulfate d'ammonium, c'est pour montrer ici une des différences structurales à laquelle nous a conduit la présence ou l'absence de sulfate d'ammonium. En effet dans la première structure, plusieurs molécules d'ion sulfate chargé négativement, en se plaçant dans le site actif, ont mimé la présence du substrat. Une molécule de coenzyme NAD^+ par monomère, a conjointement conduit la protéine à se retrouver dans sa conformation "active", c'est à dire avec la boucle catalytique refermant le site actif **[Figure 7.13b]**. L'autre structure, dépourvue de tout ligand et autre cofacteur, se montre dans sa forme "apo", boucle catalytique en conformation ouverte, site actif vide **[Figure 7.13b]**.



(a) Boucle catalytique de PtMalDH en conformation fermée. Conformation due à la présence de SO_4^{2-} dans le site actif (en surface de Connolly beige) et à la présence du cofacteur NAD^+ dans son site de liaison (en rouge)

(b) Boucle catalytique de PtMaDH en conformation ouverte correspondant à l'enzyme sous forme apo. Site actif représenté en surface de Connolly beige

Figure 7.13 – Deux conformations de la boucle catalytique du site actif du monomère A de PtMalDH. Fermée en présence de SO_4^{2-} et de NAD^+, ouverte sous forme apo.

Il convient de noter que c'est la première fois qu'il est possible d'observer par cristallographie les changements conformationnels associés à la catalyse d'une MalDH tétramérique. En effet, comme il n'existe pas d'inhibiteur compétitif de l'oxaloacétate, les réorganisations conformationnelles n'avaient été décrites jusqu'alors que pour les LacDH. Bien que ce ne soit pas illustré de manière détaillée dans ce manuscrit, la comparaison des formes ouvertes et fermées entre MalDh et LacDH montre des différences notables qui seront exploitées ultérieurement dans le cadre d'analyses phylostructurales.

7.2.3 CaMalDH, la malate déshydrogénase de Chloroflexus aurantiacus et la structure de l'eau dans son enveloppe de solvatation

La malate déshydrogénase homotétramérique de *Chloroflexus aurantiacus* est composée de 1236 acides aminés (4x309). Cela lui confère une masse molaire de 131 kDa, un point isoélectrique théorique de 5.16 et un coefficient d'extinction molaire théorique ε de 18910 $M^{-1}.cm^{-1}$. Les conditions de cristallisation ayant permis d'obtenir des cristaux diffractant à haute résolution sont décrites dans la sous-section 7.1.1 page 214.

7.2.3.1 Intégration des données utilisées pour le phasage par la méthode SIRAS

La structure de la protéine naturelle a été résolue par la méthode SIRAS à une résolution de 1,70 Å. Deux jeux de données ont été utilisés : un jeu de données de diffraction issu du cristal dérivé avec le complexe GdHPDO3A (décrit en 7.1.3.2 page 217), et celui issu d'un cristal natif enregistré au seuil d'absorption K du sélénium. Il existe déjà dans la Protein Data Bank une structure de CaMalDH à

2,20 Å de résolution (code PDB : 1GUY) mais le modèle ne comprend ni la boucle catalytique, ni les parties N-terminales et C-terminales (Dalhus *et al.*, 2002).

	CaMalDH	
Situation	Institut de Biologie Structurale	ESRF
Source de rayons X	Générateur de laboratoire	FIP-BM30A
Méthode	**SIRAS**	
Données	Jeu anomal	jeu natif
λ (Å)	1,541 ($Cu_{K\alpha}$)	0,979 (Se_K)
Groupe d'espace	$P3_121$	
Paramètres de maille (Å)	a = 106,775, c = 103,526	a = 106,230, c = 102,570
Contenu unité asymétrique	dimère A-D	
Résolution (Å)	19,63 - 1,90 (2,00 - 1,90)	19,70 - 1,70 (1,79 - 1,70)
Nombre de réflexions uniques	54120 (7772)	68673 (10178)
R_{merge} (%)	5,0 (23,7)	8,4 (35,6)
R_{pim} (%)	1,9 (8,5)	4,4 (20,3)
Sig_{ano} XDS	1,29 (0,85)	-
R_{ano} (%)	3,6 (9,6)	-
$\frac{I}{\sigma(I)}$	12,4 (3,3)	6,3 (2,2)
Complétude (%)	98,8 (99,1)	93,7 (95,8)
Redondance	10,6 (9,8)	4,2 (3,9)

Tableau 7.22 – Statistiques d'intégration des jeux de données utilisés pour le phasage de CaMalDH par la méthode SIRAS ; colonne de gauche : un cristal dérivé de CaMalDH avec le complexe GdHPDO3A ; colonne de droite : un cristal natif de CaMalDH

7.2.3.2 Phasage par la méthode SIRAS

Les statistiques de phasage par la méthode SIRAS sont répertoriées dans le tableau ci-dessous.

	CaMalDH
	SIRAS
Résolution (Å)	19,70 - 1,70
$FOM^{acen.}_{sharp}$ / $FOM^{cen.}_{sharp}$	0,235 / 0,105
Nombre de sites	8
Résolution $FOM^{acen.}_{sharp} \approx$ 0,5 (Å)	4,41
$P\varphi^{\neq isomorphes}_{sharp\,acen.}$ / $P\varphi^{\neq anomales}_{sharp\,acen.}$	1,366 / 2,672
Solvant (%)	49,5
$SIGMAA_{solomon}$	49,86
Buccaneer (construit / total)	552 / 618

Tableau 7.23 – Statistiques du phasage SIRAS de CaMalDH avec le complexe GdHPDO3A

7.2.3.3 Affinement de la structure de CaMalDH

L'affinement a été effectué avec le programme Phenix selon le protocole indiqué dans l'introduction de ce manuscrit. Les statistiques d'affinement sont consignées dans le tableau suivant.

Le dimère A-D étant le seul contenu dans l'unité asymétrique, la reconstruction du tétramère se fait par opération de symétrie cristalline. Le monomère A devenant alors B et le monomère D devenant alors C.

	CaMalDH
Code RCSB-PDB	4BGT
Résolution (Å)	19,70 - 1,70
R$_{work}$ (%)	17,02
R$_{free}$ (%)	21,02
Nbre de réflexion utilisées	67827
Composition (nbre d'atomes)	
Protéine	4694
H2O	946
Ions	12
Ligands	55
Res. hors Ramachandran (%)	0,30
Ecarts types (globaux)	
Liaisons$_{Longueurs}$ (Å)	0,010
Liaisons$_{Angles}$ (°)	1,292
B$_{moyen}$ protéine (Å^2)	21,50
B$_{min}$ protéine (Å^2)	9,17
B$_{max}$ protéine (Å^2)	109,92
B$_{moyen}$ eau (Å^2)	36,04
B$_{moyen}$ ions divers (Å^2)	34,55
B$_{moyen}$ ligands divers (Å^2)	32,47

Tableau 7.24 – Résumé des statistiques d'affinement de la structure native de CaMalDH

Je précise ici que le modèle de CaMalDH répondant au code PDB 1GUY contient une erreur : en position 34, toutes les séquences primaires disponibles dans les banques de données indiquent une phénylalanine et non pas une isoleucine.

7.2.3.4 Structuration de l'eau à la surface de CaMalDH et adaptation halophile

S'il m'est permis de raconter encore une courte histoire, je compterai comment l'étude de la structure des molécules d'eau à la surface de CaMalDH a émergé au sein de l'équipe.

Alors que Richard Kahn et moi-même étions en train de parcourir la carte de densité électronique améliorée par aplatissement de solvant, nous esclaffant devant un phasage SIRAS d'une telle précision, je tombais sur une densité un peu particulière qui nous laissa pantois. A l'écran, nous observions cinq densités électroniques de forme parfaitement sphérique, bien distinctes de celle correspondant à celles de la protéine, respectant scrupuleusement la géométrie pentagonale. A ma question : "Mais là, qu'est-ce-que je modélise là-dedans ?", mon directeur de thèse resta, comme moi, un peu perplexe. La réponse vint rapidement, au détour du second tour d'affinement, lorsque la construction automatique des molécules d'eau fut terminée.

Un grand nombre de molécules d'eau dans l'enveloppe de solvatation de CaMalDH

La première information qui saute au yeux lorsqu'on regarde la structure de la CaMalDH native est la forte hydratation de cette dernière. En fin d'affinement, le modèle contient 1,56 molécules d'eau par résidu. En 2004, Clemens Vonrhein a présenté les statistiques, basées sur tous les fichiers de la Protein Data Bank concernant le nombre de molécules d'eau par acide aminé. Cette étude a montré que pour l'ensemble des structures à 1.7 Å de résolution, il y avait 0,13 molécules d'eau par atome, ce qui fait donc environ 1,3 molécules d'eau par acide aminé (Wood *et al.*, 2008). Avec 1,56 molécules d'eau par acide aminé, nous avons donc résolu la structure d'une macromolécule biologique particulièrement hydratée **[Figures 7.14]**. Cette même figure montre aussi la taille des canaux de solvant dans les cristaux obtenus (caractéristique que nous avions exploitée dans l'étude du complexe LnTNTPA présentée dans le dernier chapitre de la première partie de ce manuscrit)..

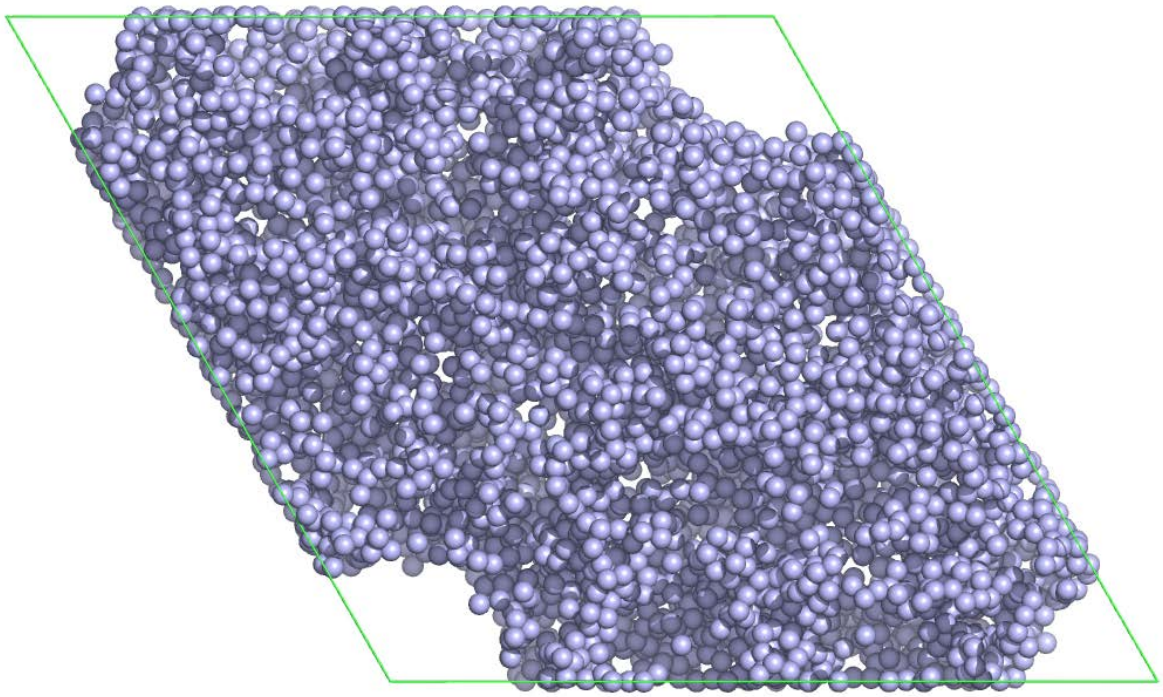


Figure 7.14 – Vue de la maille cristalline d'un cristal natif de CaMalDH. Seules les molécules d'eau sont représentées en bleu-gris

Un réseau de polygones d'eau dans la première couche de solvatation

Une observation rapide des molécules d'eau modélisées permet de constater une structuration particulière de ces dernières et ce, uniquement dans la première couche de solvatation autour de la surface de la protéine.

Les molécules d'eau coordonnées entre elles s'arrangent à la surface de la protéine CaMalDH selon trois structurations possibles : une géométrie pentagonale, une géométrie hexagonale ou en vrac. Nous avons cartographié les molécules d'eau dont la coordination adopte une géométrie pentagonale ou hexagonale. La liste exhaustive de ces molécules d'eau et leur caractéristiques structurales se trouvent dans les **tableaux 7.25 et 7.26**.

Plg	Molécules d'eau et acides aminés						Réseau	Mer	Res. Couvert	Coord. par C=O des. res.		Distances (Å)				Angles (°)			S (≥ e⁻/Å³)	Tordu
												Du res. couv.	D. min	D. max	Av. D	A. min	A. max	Av. A		
1	32	136	158	243	484		1	A	A 118 VAL	A 119 ASN	A 143 GLN	3,36	2,62	2,87	2,72	100,90	114,60	106,43	3,20	NO
2	74	136	243	298	435		1	A	A 229 PRO	A 225 ALA		3,40	2,66	3,52	2,99	78,60	112,55	100,00	3,30	YES
3	55	63	91	113	233		2	A	A 205 ILE	A 155 PHE		3,18	2,65	3,27	2,82	84,92	121,80	106,92	3,90	NO
4	16	63	91	454	926		2	A		A 155 PHE			2,58	3,87	2,96	84,15	119,05	103,37	3,80	NO
5	262	426	458	631	691		3	A	A 290 ALA	A 286 GLU		3,11	2,66	3,41	2,89	94,01	119,12	106,20	2,20	NO
6	59	286	455	458	674		3	A	A 293 ASN (lc)			3,29	2,75	3,21	2,88	83,15	123,65	105,47	1,80	YES
7	59	285	286	515	956		3	A	A 294 ALA	A 290 ALA		2,78	2,68	3,03	2,77	90,50	121,01	107,02	2,20	NO
8	61	170	297	352	588		4	A	A 203 ALA	A 203 ALA		3,12	2,61	3,00	2,82	101,49	111,46	107,74	2,40	NO
9	297	320	358	588	638		4	A	A 203 ALA	A 200 ASP		3,11	2,53	3,07	2,88	100,97	117,63	107,99	1,80	NO
10	61	212	231	274	384		4	A	A 206 VAL			4,00	2,56	3,11	2,77	87,03	114,54	101,67	3,60	YES
11	61	170	231	235	526		4	A	A 206 VAL	A 203 ALA		3,28	2,57	3,14	2,78	100,15	112,65	107,06	3,40	NO
12	231	526	533	546	747		4	A	A 291 LEU			3,40	2,56	2,95	2,72	81,58	138,38	107,67	1,80	YES
13	117	390	546	590	747		4	A	A 291 LEU	A 287 GLU		3,31	2,66	3,09	2,85	97,21	114,16	106,57	1,90	NO
14	163	167	231	274	390	546	4	A	A 291 LEU	A 181 PRO		2,94	2,56	2,99	2,81	92,40	130,57	116,32	2,50	YES
15	93	177	372	853	X		5	A	A 219 LEU	A 219 LEU		3,28	2,21	3,57	2,88				2,00	NO
16	853	1433	1505	1676	X		5	A	A 219 LEU			3,09	2,11	2,95	2,73				1,60	NO
17	210	267	362	475	532		6	A	A 101 ALA	A 65 TYR (OH)		3,32	2,65	2,88	2,75	103,13	114,83	107,80	1,70	NO
18	362	383	532	817	831		6	A	A 101 ALA			3,40	2,53	4,19	2,96	83,79	113,93	102,69	1,30	YES
19	301	397	474	945	X		7	A	A 34 ILE	A 34 ILE		3,43	2,64	3,11	2,85				1,50	NO
20	115	301	404	423	A 63 ASN		7	A	A 65 TYR			3,50	2,67	3,75	2,92	91,83	124,68	107,70	2,20	NO
21	465	577	910	1820	X		8	A	B 219 LEU			2,98	2,18	2,95	2,74				1,60	NO
22	1063	1147	1341	1820	X		8	A	B 219 LEU	B 215 GLU		3,24	2,28	3,68	2,91				2,00	NO
23	277	322	540	591	664			A	A 109 LEU	A 109 LEU	A 108 PRO	3,43	2,68	3,03	2,77	93,87	120,04	105,76	2,20	NO
24	71	268	693	923	A 133 SER			A	A 135 PHE			3,56	2,61	3,09	2,92	85,05	121,01	102,92	2,20	YES
25	64	300	518	930	967			A	A 207 GLU (lc)	A 204 GLN		2,84	2,60	3,88	3,09	79,08	121,82	105,96	1,90	NO
26	40	41	436	632	X			A	A 291 LEU	A 179 MET		2,80	2,51	2,76	2,67	107,00	109,46	108,24	4,20	NO
27	70	209	272	385	A 2 ARG			A	A 4 LYS (lc)	A 70 ASN		3,30	2,65	3,13	2,82	94,98	117,67	106,61	2,50	NO
28	107	172	254	469	A 112 ASN	A 273 ALA		A					2,65	2,98	2,80	87,39	127,80	111,60	2,50	YES
29	114	245	336	845	890			A					2,52	2,96	2,81	86,10	115,72	102,74	2,50	YES
30	405	1347	1456	1502	1769			A		A 52 ILE	C 165 GLU		2,63	3,02	2,83	94,36	117,40	107,32	1,80	NO
31	21	29	36	110	125		9	D	D 118 VAL	D 119 ASN	D 143 GLN	3,30	2,63	2,85	2,74	99,73	114,31	106,77	4,00	NO
32	29	66	86	123	275		9	D	D 13 VAL	D 118 VAL	D 225 VAL	2,92	2,66	3,34	2,88	88,92	120,41	107,36	3,80	NO
33	21	29	66	68	219		9	D	D 229 PRO	D 225 ALA		3,46	2,73	3,65	3,01	80,44	110,93	100,00	4,10	YES
34	29	66	125	456	536		9	D					2,58	2,77	2,69	92,17	121,40	105,64	3,20	NO

Tableau 7.25 – Cartographie des molécules d'eau à la surface du dimère A-D de CaMalDH, 1èrepartie.

Plg	Molécules d'eau et acides aminés						Réseau	Mer	Res. Couvert	Coord. par C=O des. res.		Distances (Å)				Angles (°)			S (≥ e⁻/Å³)	Tordu
												Du res. couv.	D. min	D. max	Av. D	A. min	A. max	Av. A		
35	39	238	313	1152 sym	D 197 ILE		10	D	D 198 ALA			2,58	2,58	3,47	2,87	94,42	113,15	108,97	4,10	YES
36	38	194	441	452	517		10	D	D 199 PRO			3,14	2,57	3,13	2,88	93,69	106,38	100,27	2,90	YES
37	347	452	517	621	627		10	D	D 199 PRO	C 276 VAL		2,89	2,55	3,81	2,96	84,19	113,92	104,89	1,90	NO
38	238	313	347	398	621		10	D	D 199 PRO	D 197 ILE		2,86	2,68	3,81	3,00	87,54	117,58	106,60	2,40	YES
39	94	194	398	517	621		10	D	D 199 PRO			3,29	2,58	3,13	2,82	103,30	113,03	107,38	2,40	NO
40	46	94	175	398	941		10	D	D 202 LEU			3,18	2,67	4,16	3,21	77,66	113,88	113,84	2,80	YES
41	73	76	90	325	453		10	D	D 203 ALA	D 203 ALA		3,04	2,74	3,01	2,82	99,42	116,90	107,69	2,40	NO
42	90	169	410	453	481		10	D	D 203 ALA	D 200 ASP		3,00	2,63	3,01	2,81	98,49	117,87	107,93	2,00	NO
43	38	90	194	325	481		10	D	D 203 ALA			3,20	2,63	3,42	2,86	99,04	111,93	106,92	2,40	NO
44	46	76	94	194	325		10	D	D 203 ALA	D 199 PRO		3,31	2,67	3,42	2,87	98,38	112,86	105,94	3,40	NO
45	46	76	128	175	445		10	D	D 206 VAL			3,95	2,61	2,86	2,75	86,72	111,35	99,30	3,60	YES
46	67	73	76	88	128		10	D	D 206 VAL	D 203 ALA		3,28	2,58	2,91	2,70	97,18	116,20	107,04	3,60	NO
47	35	197	224	357	757		10	D	D 288 GLU (lc)	D 181 PRO		3,02	2,48	3,39	2,81	96,44	115,70	106,15	2,50	YES
48	357	447	519	757	881		10	D					2,53	3,32	2,84	82,90	115,12	103,63	1,50	YES
49	175	197	224	357	445	447	10	D		D 181 PRO			2,48	3,32	2,81	95,62	113,79	102,27	2,20	YES
50	322	683	1153	1187	1307	1362	11	D	C 219 LEU			3,15	2,39	2,93	2,67	95,40	124,62	110,15	2,00	YES
51	918	1072	1187	1307	1530		11	D	C 219 LEU	C 215 GLU		3,19	2,59	3,01	2,79	84,25	132,40	106,39	2,30	NO
52	166	322	391	918	1307		11	D	C 219 LEU			3,06	2,55	3,01	2,74	98,29	122,52	107,84	1,70	NO
53	199	241	683	790	1362		11	D	D 40 GLN (lc)	D 40 GLN	D 61 GLY	3,40	2,54	3,74	2,88	97,69	113,52	107,08	2,00	NO
54	60	199	241	842	920		11	D	D 60 THR (ch3)	D 40 GLN		3,73	2,54	3,08	2,80	84,67	115,87	99,68	2,10	YES
55	535	615	783	784	902	920	11	D	D 60 THR (ch3)			2,97	2,56	3,00	2,82	81,45	133,24	112,00	1,70	YES
56	393	1169	1211	1651	1757		12	D	C 40 GLN (lc)	C 40 GLN	C 61 GLY	3,28	2,54	3,74	2,88	97,68	115,52	107,08	2,00	NO
57	102	217	338	562	1884		12	D	D 219 LEU	D 215 GLU		3,22	2,59	3,01	2,80	84,24	132,41	106,39	2,30	NO
58	183	217	338	393	1291	1651	12	D	D 219 LEU			3,03	2,39	2,93	2,67	93,41	124,61	110,15	2,00	YES
59	338	1136	1291	1360	1884		12	D	D 219 LEU			3,22	2,55	3,01	2,75	98,32	122,50	107,84	1,70	NO
60	124	140	216	438	462		13	D	A 270 ILE	D 160 ALA	D 161 GLY	3,78	2,73	3,47	2,98	99,60	112,64	106,95	2,90	NO
61	216	363	414	438	D 160 ALA		13	D	A 270 ILE			3,54	2,58	2,99	2,76	89,61	124,58	107,20	2,90	NO
62	260	318	319	411	479	878	14	D	D 291 LEU			2,79	2,57	2,93	2,71	107,55	127,66	118,88	1,60	NO
63	206	247	480	549	658		14	D	D 294 ALA	D 290 ALA		3,05	2,69	3,09	2,85	98,02	115,38	107,56	1,70	NO
64	288	480	549	731	788		14	D	D 294 ALA	D 294 ALA		3,14	3,03	4,08	3,35	87,78	123,78	108,05	1,80	NO
65	318	479	480	658	788		14	D	D 294 ALA			3,13	2,69	3,14	2,87	101,73	116,16	207,09	1,50	NO
66	30	202	318	731	788		14	D	D 294 ALA	D 177 ASP	D 291 LEU	3,03	2,72	4,08	3,04	94,55	117,18	107,60	2,90	NO
67	206	247	596	662	714		14	D		D 290 ALA			2,71	3,73	3,02	73,86	138,74	105,87	1,80	YES
68	30	104	202	260	318	D 177 ASP	14	D					2,61	2,93	2,78	90,93	141,46	107,24	3,00	YES
69	242	261	324	494	572		15	D	D 101 ALA	D 65 TYR (OH)		3,20	2,61	2,84	2,78	103,10	111,95	107,22	2,00	NO
70	494	572	635	886	X		15	D	D 101 ALA			3,40	2,48	3,41	2,83				1,70	NO
71	1	20	176	376	D 112 ASN	D 273 ALA		D	D 114 VAL	D 112 ASN	D 273 ALA	4,02	2,68	2,89	2,77	87,65	131,05	113,23	4,20	YES
72	179	190	265	334	403			D	D 208 ARG (lc)	D 204 GLN		3,40	2,58	2,88	2,73	85,29	123,03	106,60	2,70	NO
73	152	162	184	259	D 247 VAL			D	D 248 MET (lc)			3,14	2,63	3,60	2,93	85,74	119,17	103,52	3,00	YES
74	24	50	75	1124	D 25 GLU			D	D 26 LEU			2,77	2,63	3,57	2,96	88,29	106,73	100,17	3,40	YES
75	82	186	335	496	D 2 ARG			D	D 4 LYS (lc)	D 70 ASN		3,46	2,63	3,23	2,91	100,53	112,61	105,59	3,20	NO
76	14	120	198	215	430	704		D					2,64	3,04	2,76	91,08	142,52	115,38	2,40	NO

Tableau 7.26 – Cartographie des molécules d'eau à la surface du dimère A-D de CaMalDH, $2^{ème}$ partie

Abréviations utilisées dans **les tableaux 7.25 et 7.26**.

Réseau Numéro d'un réseau de molécules d'eau constitué d'au moins deux polygones,

Plg Numéro du polygone,

Mer Monomère de la protéine correspondant.

Res. Couvert Acide aminé couvert par le polygone. (**lc**) correspond à "chaine latérale", c.a.d "chaine aliphatique d'un grand acide aminé", (**ch3**) signifie que seule cette fonction chimique hydrophobe est recouverte par le polygone.

Coord par C=O des. res. Ici sont énumérés les acides aminés dont l'oxygène de la fonction peptidique coordonne le polygone. (**OH**) est là pour préciser qu'exceptionnellement, c'est la fonction hydroxyle de la tyrosine qui coordonne le polygone.

Distances Du. res. Couv. Distance la plus courte mesurée entre l'atome le plus haut de la chaine latérale de l'acide aminé couvert et le plan formé par le polygone.

D. min Distance la plus courte mesurée entre deux molécules d'eau qui constituent le périmètre du polygone.

D. max Distance la plus longue mesurée entre deux molécules d'eau qui constituent le périmètre du polygone.

Av. D Moyenne des distances mesurées entre deux molécules d'eau du périmètre du polygone.

A. min Angle le plus petit mesuré entre trois molécules d'eau qui constituent le périmètre du polygone.

A. max Angle le plus grand mesuré entre trois molécules d'eau qui constituent le périmètre du polygone.

Av. A Moyenne des angles mesurés entre trois molécules d'eau du périmètre du polygone.

S ($\geq e^-/Å^3$) Valeur de sigma nécessaire pour faire disparaître la densité électronique d'au moins trois molécules d'eau du polygone. Valeur qui peut être interprétée comme la probabilité d'existence du polygone. Plus cette valeur est élevée, plus l'existence de ce polygone à la surface de la protéine est probable.

Tordu Le polygone s'est tordu pour s'adapter à la surface de la protéine ou non.

Qu'elle que soit l'organisation, les distances moyennes permettant une coordination par liaison hydrogène entre les molécules d'eau ou avec les acides aminés de la surface protéique sont toujours respectées : environ 2,80 +/-0,06 Å. Il en est de même pour les angles entre trois molécules d'eau. L'angle est formé par l'axe d'une liaison hydrogène entre l'atome d'hydrogène d'une première molécule d'eau avec le doublet non-liant de l'atome d'oxygène d'une deuxième et l'axe de la liaison hydrogène entre l'atome d'hydrogène de la deuxième et le doublet non-liant de l'atome d'oxygène d'une troisième molécule d'eau. La valeur moyenne de ces angles est de 107 ° (Weik and Colletier, 2010) **[Tableau 7.27]**.

Dans le cas de réseaux de pentagones, ces valeurs peuvent fluctuer légèrement autour de la valeur moyenne mais l'ensemble du réseau, en s'étirant ou en se comprimant, maintient un ensemble énergiquement favorable **[Tableau 7.27]**.

De plus, bien que les polygones soient plans dans 64,5 % des cas, on constate qu'ils peuvent se tordre pour s'adapter à des contraintes structurales de surface. C'est par exemple le cas lorsque un pentagone recouvre partiellement une proline. Par rapport au plan formé par les quatre autres molécules d'eau, la pointe du pentagone la plus proche de la proline se relève permettant de maintenir une distance d'environ 3,20 +/-0,8 Å entre le pentagone et la surface de la protéine **[Tableau 7.27]**.

CaMalDH Dimère A-D		
Nombre de molécules d'eau total :	972	(x2 pour le tetramère complet) (Chaine B = A et chaine C = D)
Nombre de molécules d'eau par résidus :	1,57	
Molécules d'eau constituant des polygones:	28,09%	(273 / 932)
Nombre de polygones trouvés :	76	
Nombre de pentagones trouvés :	66	
Nombre d'hexagones trouvés :	10	
Nombre de réseaux de polygones différents :	15	
Plus grand réseau de polygones trouvé :	15 polygones associés	
Pourcentage de polygones tout à fait plans :	64,47%	(49 / 76)
Distance minimale observée entre le résidu couvert et le plan formé par un polygone :		2,58 Å
Distance maximale observée entre le résidu couvert et le plan formé par un polygone :		4,02 Å
Distance moyenne entre les résidus couverts et les plans formés par les polygones :		3,23 Å
Distance minimale observée entre deux molécules d'eau du périmètre d'un pentagone :		2,11 Å
Distance maximale observée entre deux molécules d'eau du périmètre d'un pentagone :		4,19 Å
Distance moyenne entre deux molécules d'eau du périmètre des pentagones :		2,86 Å
Distance minimale observée entre deux molécules d'eau du périmètre d'un hexagone :		2,39 Å
Distance maximale observée entre deux molécules d'eau du périmètre d'un hexagone :		3,32 Å
Distance moyenne entre deux molécules d'eau du périmètre des hexagones :		2,76 Å
Angle minimal observé entre trois molécules d'eau du périmètre d'un pentagone :		73,86 °
Angle maximal observé entre trois molécules d'eau du périmètre d'un pentagone :		138,74 °
Angle moyen observé entre les molécules d'eau du périmètre des pentagones :		107,56 °
Angle minimal observé entre trois molécules d'eau du périmètre d'un hexagone :		81,45 °
Angle maximal observé entre trois molécules d'eau du périmètre d'un hexagone :		142,52 °
Angle moyen observé entre les molécules d'eau du périmètre des hexagones :		111,72 °

Tableau 7.27 – Résumé des caractéristiques structurales des molécules d'eau à la surface de CaMalDH

La caractéristique principale de ces réseaux de polygones d'eau est leur situation dans la première couche de solvatation "recouvrant" exclusivement les acides aminés apolaires se trouvant en surface (glycine, alanine, valine, leucine, isoleucine, proline, phénylalanine et tryptophane). Il est à noter que certains polygones forment des superstructures **[Figure 7.15]**.

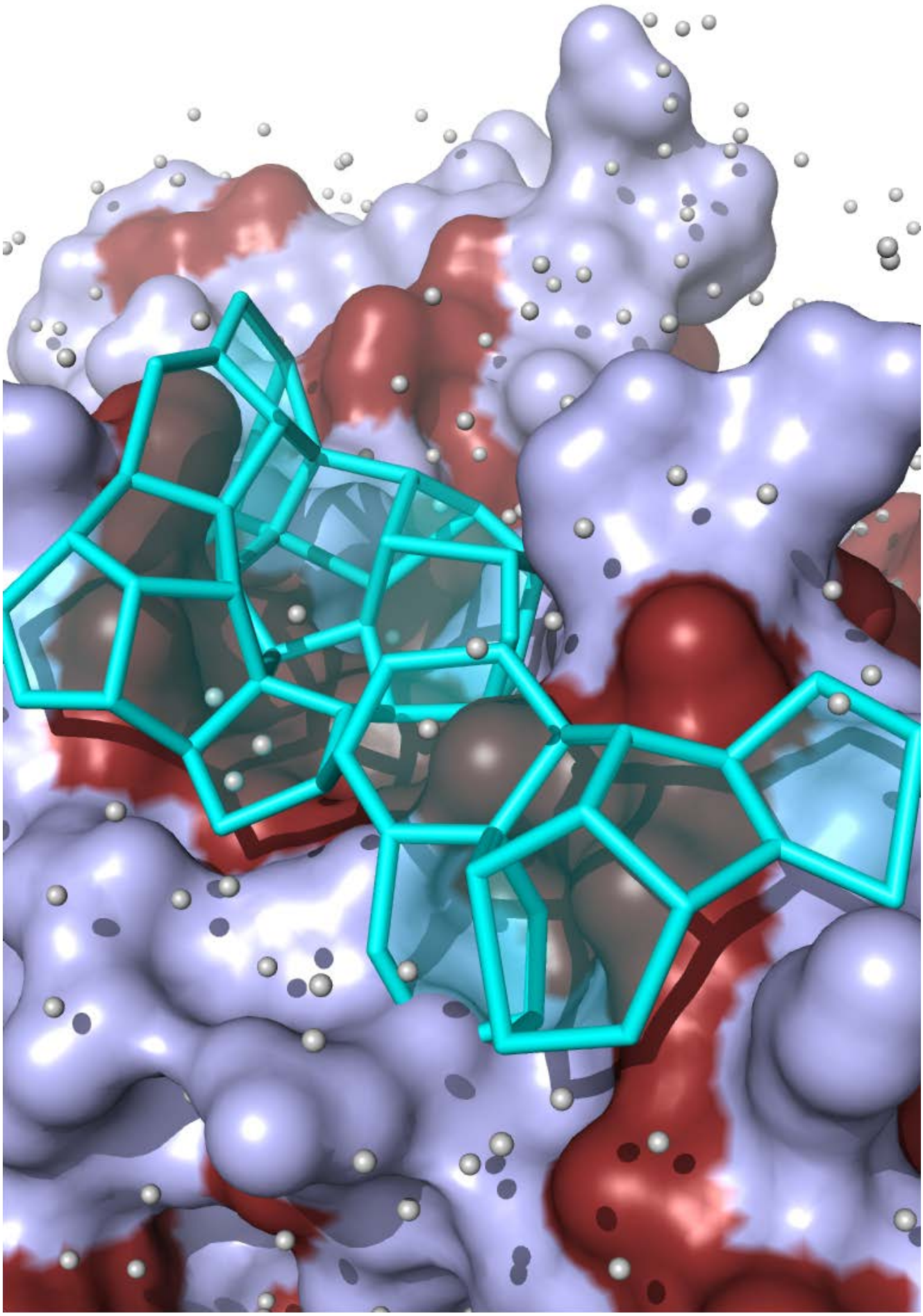


Figure 7.15 – Réseau de polygones d'eau à la surface de CaMalDH. En bleu-gris, la surface de Connolly des acides aminés hydrophiles, en chocolat, celle des acides aminés hydrophobes

Dès qu'un acide aminé polaire est présent à la surface, le réseau s'interrompt pour laisser la place à la chaine latérale hydrophile qui coordonne alors les molécules d'eau non-impliquées dans un polygone **[Figure 7.16]**.

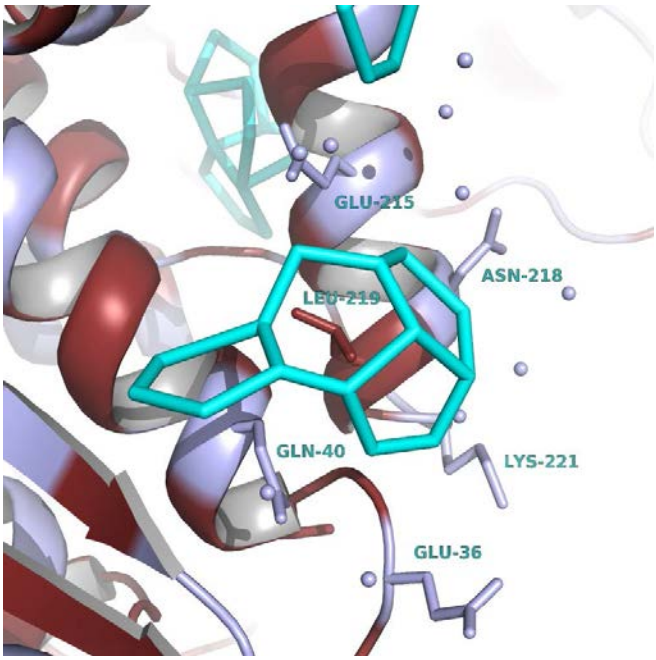


Figure 7.16 – Réseau de pentagones d'eau recouvrant une leucine de la surface de CaMalDH exposée au solvant. Couleur chocolat, les acides aminés hydrophobes, en bleu gris, les acides aminés hydrophiles dont les chaines latérales empêchent le réseau de polygones de s'étendre

Plus précisément, les polygones recouvrent les chaines latérales hydrophobes des acides aminés qui constituent la surface protéique. Cela signifie que, si la forme de la zone le permet, il est possible de trouver, non seulement des pentagones dont le plan se situe précisément à la perpendiculaire de la fonction méthyle de la chaine latérale d'une thréonine mais aussi des pentagones dont le plan est parallèle à la chaine aliphatique qui constitue la chaine latérale d'une lysine ou d'une arginine. Or, ces trois acides aminés sont pourtant polaires. Dans ces trois cas, le périmètre du pentagone ne surplombe jamais le groupement polaire (respectivement hydroxyle, guanidinium et amine) de ces acides aminés **[Figure 7.17]**.

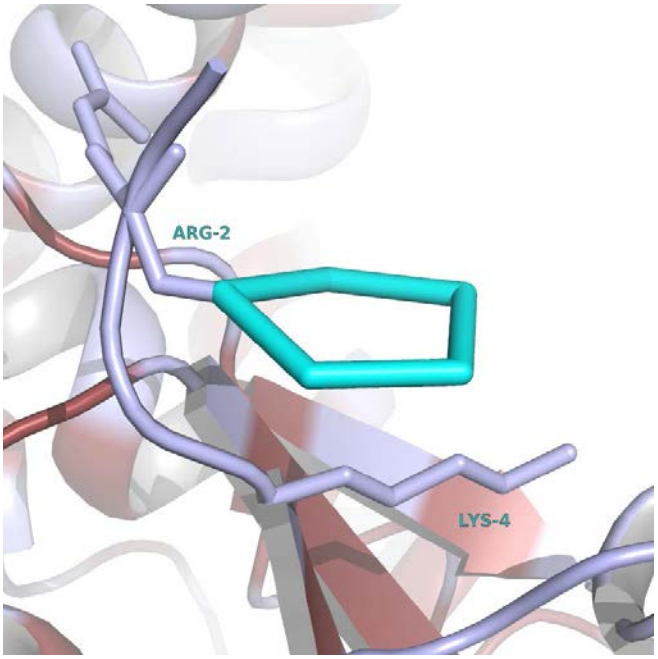


Figure 7.17 – Polygone d'eau à la surface de CaMalDH parallèle à la chaine aliphatique d'une lysine. L'oxygène de la liaison petptidique d'un acide en amont fait partie intégrante du pentagone

Les polygones sont exclusivement coordonnés avec la surface protéique par l'atome d'oxygène des liaisons peptidiques. Dans un petit nombre de cas, cet atome d'oxygène fait partie intégrante des pentagones **[Figure 7.17]**. Je n'ai constaté qu'une seule fois la coordination d'un pentagone par l'atome d'oxygène du groupement hydroxyle d'une tyrosine.

Les **Figures 7.18 et 7.19** sont deux exemples de ce que j'ai pu observer et que je considère comme des polygones d'eau "structurants". Le premier exemple **[Figures 7.18]** montre la participation de certains pentagones dans la structure tertiaire de la protéine. Le second exemple **[7.19]** concerne les réseaux de pentagones présents autour et dans la poche des sites catalytiques de CaMalDH.

Les polygones d'eau à la surface de CaMalDH ont de commun avec les chanterelles "Trompette des morts" (*Craterellus cornucopioides*, Pearson, 1825) la méthode par laquelle on les trouve : il est bien plus probable d'en trouver dans une zone moins hydratée et lorsqu'on en trouve un, en soulevant un peu l'humus, de proche en proche, on en trouve mille.

En effet, à la surface de CaMalDH, plus de 28 % des molécules d'eau modélisées sont impliquées dans des polygones. 76 polygones, dont 66 pentagones et 10 hexagones ont été observés. 15 réseaux de polygones distincts ont pu être construits et le plus étendu de ces réseaux associe 15 polygones.

Enfin, je n'ai jamais constaté la présence de polygones de molécules d'eau ni dans la cavité centrale de CaMalDH, ni aux interfaces monomère-monomère : tous les polygones recensés se situent à la surface de la protéine. Ceci peut s'expliquer par le fait que, d'une part, la surface de la cavité centrale est riche en acides aminés polaires et que, d'autre part, il y a un encombrement stérique important aux interfaces monomère-monomère qui ne laisse pas la place à la formation de polygones de molécules d'eau. Pour l'instant, suite à ces premières observations, la structuration de l'eau en polygone ne semble donc pas intervenir dans la structure quaternaire de la protéine. Ce qui m'amène à discuter du rôle de la structuration de l'eau dans l'adaptation halophile.

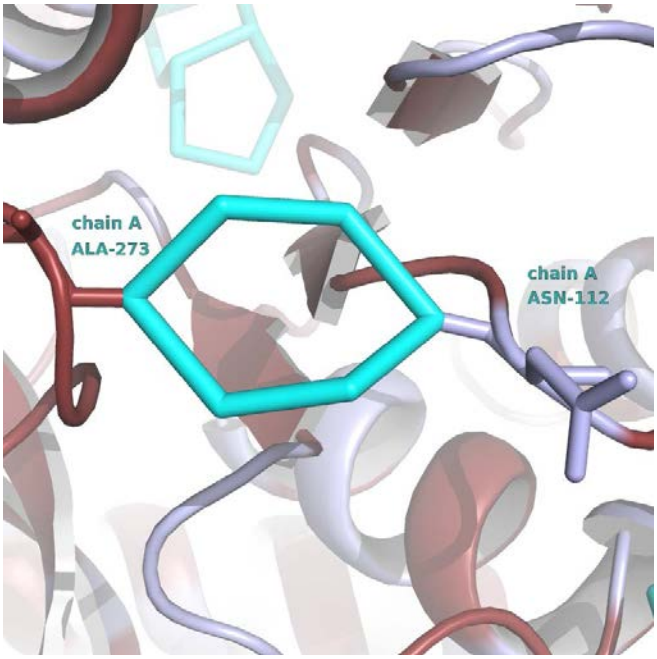


Figure 7.18 – Polygone d'eau "structurant" entre les oxygènes des liaisons peptidiques de deux acides aminés non-adjacents

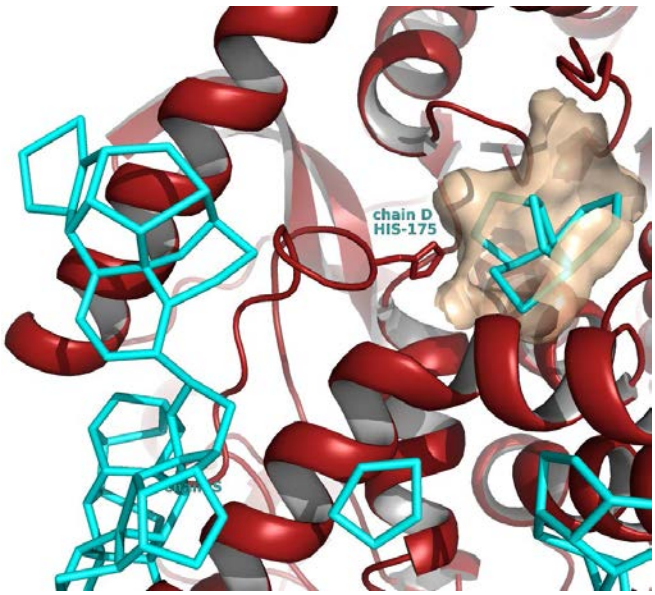


Figure 7.19 – Réseau de polygones d'eau à proximité et dans le site actif du monomère D de CaMalDH. La poche du site actif est représentée en surface de Connolly de couleur beige, le monomère D est entièrement coloré en rouge

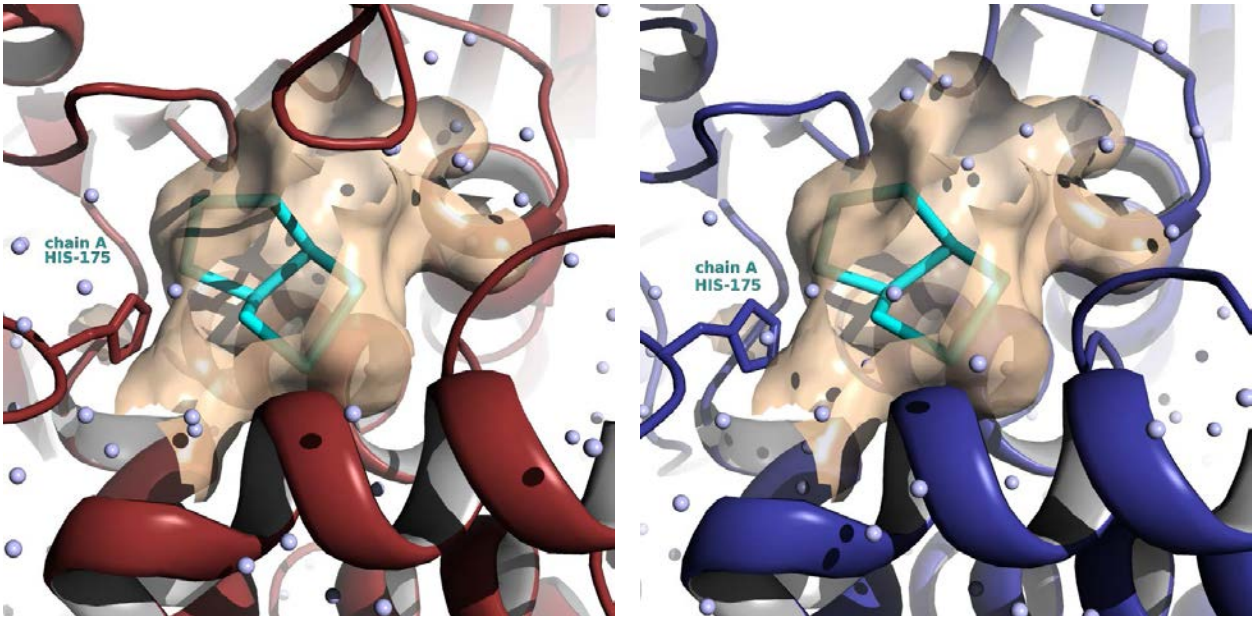
Rôle de la structure des molécules d'eau dans l'adaptation halophile

Dans le but d'apporter un nouvel éclairage sur l'adaptation halophile, deux études sont menées de conserve au groupe E.L.M.A : l'étude des substitutions adaptatives de surface chez les protéines halophiles et non-halophiles et l'étude de la structuration de l'eau à leur surface qui en découle.

Nous avons comparé la structure de CaMalDH (non-halophile) présentée dans ce chapitre, avec celle de la malate déshydrogénase de Salinibacter ruber "SrMalDH" (halophile) résolue par le Dr. Nicolas

Coquelle, ancien étudiant en thèse du Dr. Madern et actuellement en post-doctorat à l'Université d'Alberta à Edmonton (Canada). Comme nous l'avons vu dans l'introduction, ces deux protéines partagent plus de 72 % de similarité de séquences mais la surface de SrMalDH est nettement plus polaire, on ne devrait donc quasiment plus retrouver de réseau de polygones de molécules d'eau à sa surface.

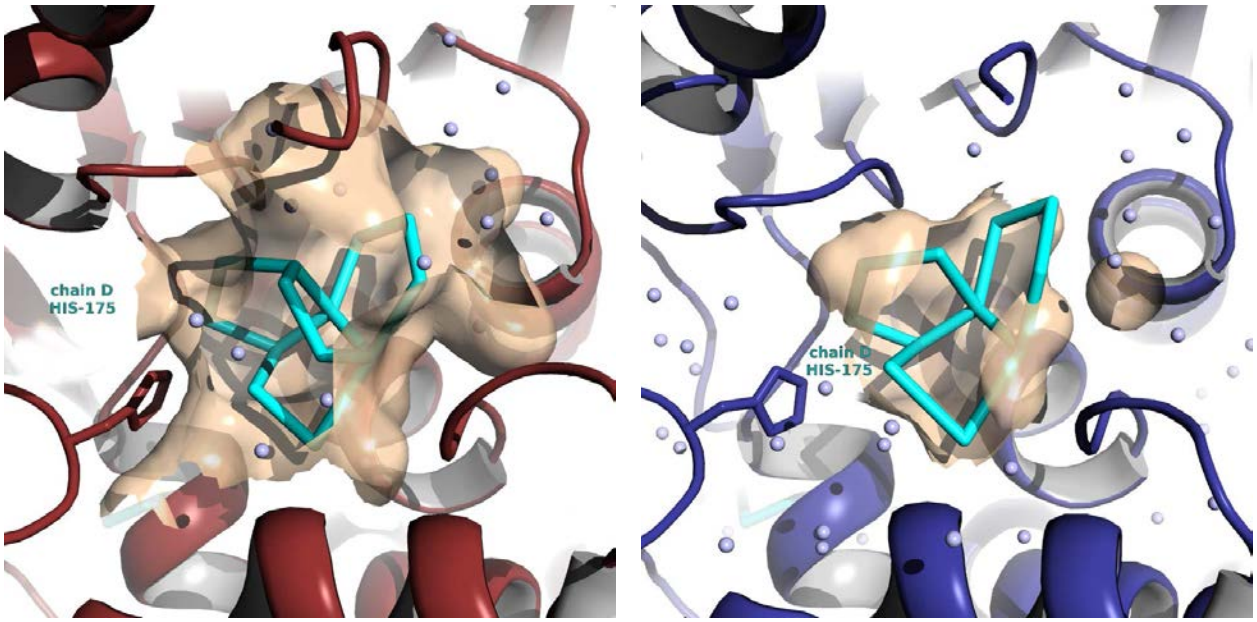
Les premières observations issues de la comparaison de ces deux structures montrent que seuls trois réseaux de polygones d'eau se retrouvent à la surface de SrMalDH. Ces réseaux se retrouvent chez les deux enzymes dans des régions très conservées : dans les deux sites catalytiques hydrophobes **[Figure 7.20 et 7.21]** et dans la région de la leucine 241, à l'interface du monomère D et C **[Figure 7.22]**. Nous avons pris le parti de mentionner ce dernier exemple malgré le fait que certaines molécules d'eau constituant le réseau de pentagone n'aient pas été construites dans le modèle SrMalDH. Lorsqu'on superpose les deux structures, la forme générale du réseau apparait de façon évidente. Nous pensons que l'absence de molécule d'eau est ici du à un protocole d'affinement utilisé qui diffère du notre. Dans la **figure 7.21**, ce sont les différentes conformations des boucles autour de la poche catalytique qui réduisent la taille de celle du monomère D chez SrMalDH.



(a) Réseau de polygones dans la poche du site actif (surface de Connolly beige) de la chaine A de CaMalDH (bordeau)

(b) Réseau de polygones dans la poche du site actif (surface de Connolly beige) de la chaine A de SrMalDH (bleu)

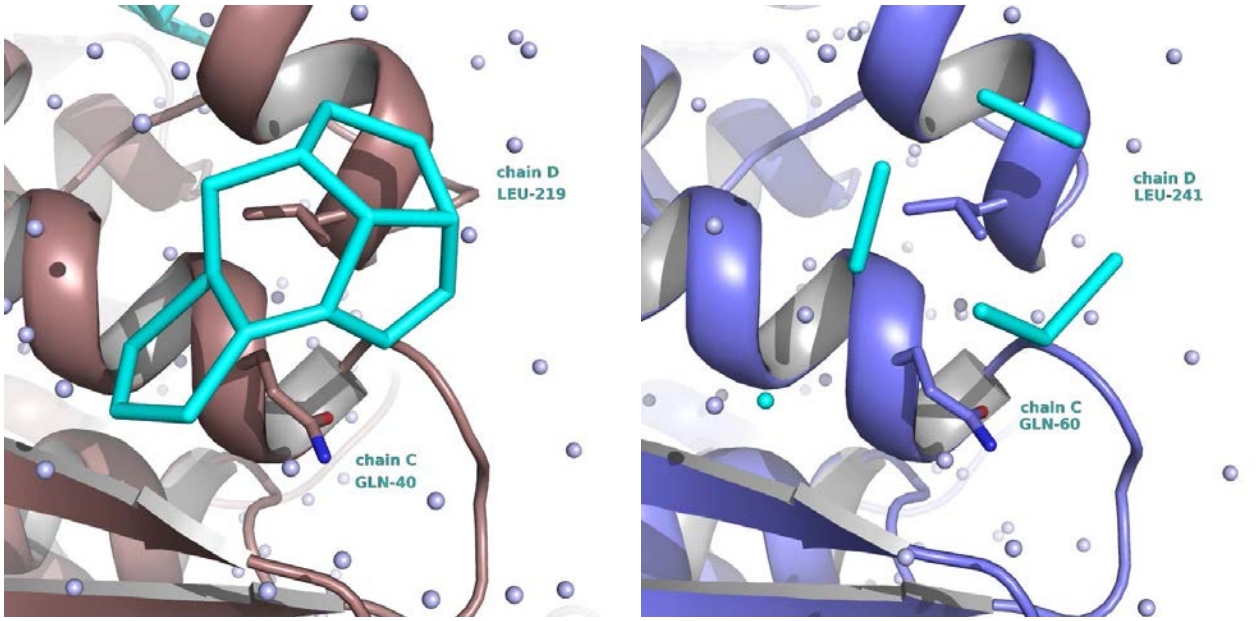
Figure 7.20 – Réseau de polygones conservé dans le site actif de la chaine A de CaMalDH (non-halophile) et de SrMalDH (halophile)



(a) Réseau de polygones dans la poche du site actif (surface de Connolly beige) de la chaine D de CaMalDH (bordeau)

(b) Réseau de polygones dans la poche du site actif (surface de Connolly beige) de la chaine D de SrMalDH (bleu)

Figure 7.21 – Réseau de polygone conservé dans le site actif de la chaine D de CaMalDH (non-halophile) et de SrMalDH (halophile)

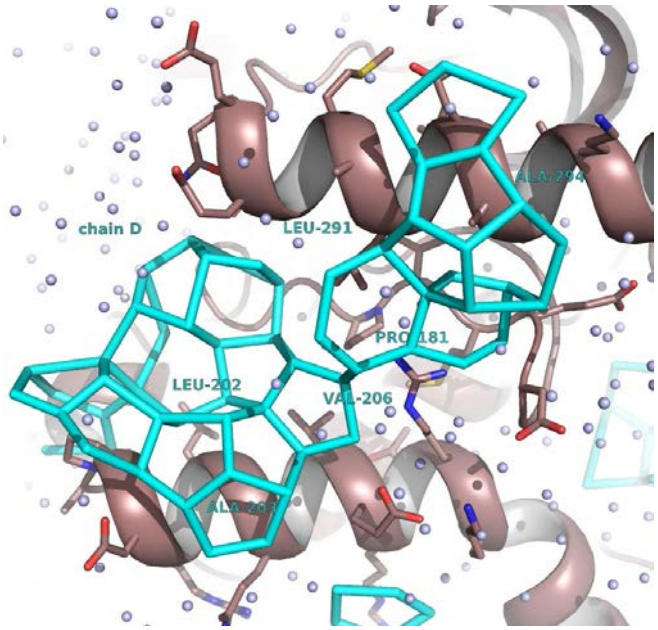


(a) Réseau de polygones recouvrant la leucine 241 à l'interface des monomères C et D de CaMalDH (bordeaux)

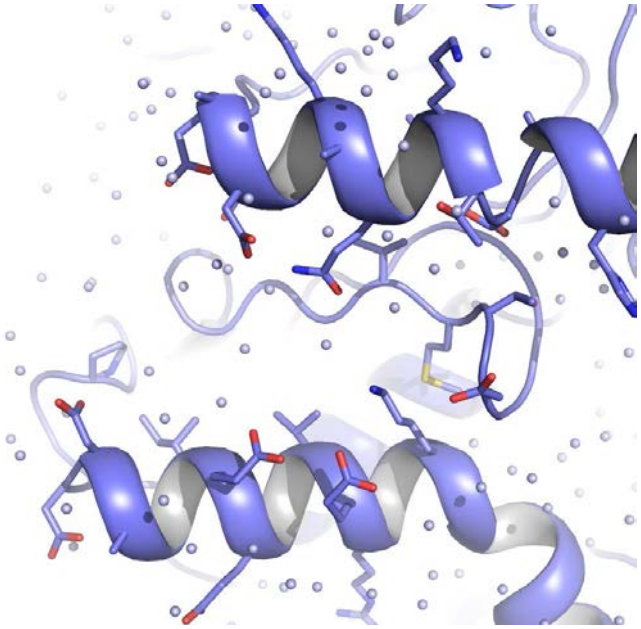
(b) Réseau de polygones recouvrant la leucine 241 à l'interface des monomères C et D de de SrMalDH (bleu)

Figure 7.22 – Réseau de polygones conservé au dessus d'un site hydrophobe conservé chez les deux enzymes CaMalDH (non-halophile) et SrMalDH (halophile)

Les exemples sont nombreux où des substitutions adaptatives d'acides aminés hydrophobes de CaMalDH en acides aminés hydrophiles chez SrMalDH perturbent l'organisation en polygone de l'eau à la surface de SrMalDH. Un seul exemple est donc mis en image dans la **figure 7.23**.



(a) Molécules d'eau en réseau de polygones recouvrant une région de la chaine D de CaMalDH (bordeaux)



(b) Molécules d'eau au dessus de la même région chez SrMalDH (bleu)

Figure 7.23 – Réseau de polygones non-conservé au dessus de la même région chez les deux enzymes CaMalDH (non-halophile) et SrMalDH (halophile)

Les premières observations issues de cette étude structurale comparative corroborent les diverses analyses en solution menées par les Drs G. Zaccai, C. Ebel et D. Madern (Madern, 2002; Coquelle *et al.*, 2010). Ces analyses ont montré que les interactions protéine-solvant sont une composante essentielle des interactions faibles protéine-protéine (Costenaro *et al.*, 2002). La différence de composition chimique entre la couche d'hydratation et le solvant « bulk » peut induire des interactions protéine-protéine attractives conduisant à la précipitation. La différence de composition entre enveloppe de solvatation et solvant doit donc être aussi "neutre" que possible : elle doit permettre des interactions protéine-

protéine répulsives afin d'assurer une grande solubilité, tout en évitant un enrichissement de la couche d'hydratation en eau ou en solvant.

Les protéines halophiles ont principalement évolué en modifiant leurs surfaces d'interaction avec le solvant, ce qui se traduit par un enrichissement en acides aminés acides en interactions avec le solvant. Ce changement de surface autorise donc une modification de l'enveloppe de solvatation des protéines halophiles, en solution par rapport à leurs homologues non halophiles.

Mes observations démontrent que la sélection à position équivalente d'un acide aminé acide perturbe la structure de l'eau dans l'enveloppe de solvatation d'une enzyme halophile par rapport à son homologue non halophile. Pour la première fois et par une expérience de cristallographie, il aura donc été possible de démontrer que la minimisation de l'exposition au solvant de groupements hydrophobes, défavorable d'un point de vue énergétique à haut sel (Thomas and Elcock, 2006), est la principale force évolutive qui dicte l'adaptation des protéines halophiles.

Les polygones d'eau dans la littérature

L'association des molécules d'eau en réseaux de polygones à la surface des protéines a déjà été observée et l'étude générale de la structure des molécules d'eau à la surface des protéines est d'actualité (Sterpone *et al.*, 2010). Le professeur Masayoshi Nakasako, de l'Université Keiö (Japon) est en particulier le le seul à avoir mis en lumière cette caractéristique structurale à travers une étude statistique poussée s'appuyant sur l'ensemble des structures de la Protein Data Bank (Matsuoka and Nakasako, 2009) profitant d'un programme de détection de polygones de son cru. Toutefois pour ses travaux, l'hypothèse de base que le Pr. Nakasako utilise est qu'un polygone de molécule d'eau peut être constitué par seulement trois molécules d'eau si les contraintes de distance et de coordination sont respectées (Nakasako, 2004; Matsuoka and Nakasako, 2009). Après avoir établi la cartographie complète des molécules d'eau à la surface de l'enzyme CaMalDH ainsi que leurs caractéristiques structurales, je pense que nous ne pouvons nous appuyer sur la même hypothèse dans le cadre de notre étude. En effet, si je m'en suis tenu à la détection de pentagones et d'hexagones de molécules d'eau dès le début de cette étude, c'est parce qu'il s'agit des deux structures géométriques qui se manifestent de façon évidente lorsqu'on regarde simplement la structure de CaMalDH, structures qui sont très favorables d'un point de vue énergétique. Par ailleurs, si trois molécules d'eau peuvent suffire pour constituer un polygone, il devient possible de créer des réseaux de polygones extrêmement étendus (Matsuoka and Nakasako, 2009). Dans le cadre de notre étude du rôle des molécules d'eau dans l'adaptation halophile, amoindrir les contraintes afin de pouvoir étendre davantage des réseaux de polygones pourrait entrainer la perte d'informations et aboutir à des conclusions moins contrastées.

7.3 Conclusions et perspectives

Les premiers résultats obtenus par la nouvelle approche méthodologique pour la compréhension fine des modes de fixation des complexes de lanthanide sont très prometteurs. L'étude d'ores et déjà menée sur trois MalDH différentes indiquent des fixations de trois complexes IRM qui diffèrent selon la composition des surfaces. Disposer d'un ensemble grandissant de conditions physico-chimiques devrait permettre d'affiner la compréhension des interactions entre la protéine et les complexes IRM. En tenant systématiquement compte de la composition des différentes surfaces des protéines modèles MalDH, de la

répartition des charges à leurs surfaces, de l'influence des conditions de cristallisation et de l'empilement cristallin (pour les cristaux dérivés par trempage) pour chacune des structures résolues, multiplié par le nombre de sites de fixation par structure, augure une étude où l'ennui n'aura pas sa place. Cela est d'autant plus vrai que l'équipe a déjà en main la production et la purification d'autre MalDH aux propriétés de surface encore différentes, comme la MalDH de l'archée alkalophile (pH basique) halophile *Natrialba magadii* ou celle de la bactérie piézothermophile *Thermoanaerobacter marianensis*. A l'issue de ce travail, nous pouvons espérer déterminer quel sera le meilleur complexe en fonction de la séquence en acide aminé de la protéine cible.

En outre, ce chapitre a de nouveau montré la force des complexes de lanthanide pour la détermination des structures par les méthodes de phasage expérimental. Ainsi, le complexe GdHPDO3A a permis de résoudre deux nouvelles structures, celle de la malate déshydrogénase de *Chloroflexus aurantiacus* et celle de la malate déshydrogénase de *Picrophilus torridus,* en permettant alors le phasage des 620 acides aminés qui composent leur dimère A-D. Le complexe GdDO3A a, lui, permis de résoudre aussi la structure de la malate déshydrogénase de *Chloroflexus aurantiacus* mais aussi celle de la malate déshydrogénase de *Haloferax volcanii*, en permettant le phasage, cette fois-ci, des 1216 acides aminés qui composent cette enzyme. Quant au complexe GdDTPA-BMA, il a, lui aussi, permit la détermination de la structure de la malate déshydrogénase de *Haloferax volcanii.*

Enfin, à travers ce chapitre, nous avons aussi illustré ce que la détermination de ces trois structures a apporté pour la compréhension de l'adaptation halophile : l'identification et le positionnement d'ions dans la structure de HvMalDH, les conformations de la boucle catalytique du site actif de HvMalDH en lien avec l'activité enzymatique et l'organisation singulière de l'eau à la surface d'une malate déshydrogénase non-halophile et d'une autre halophile. Comparer plus finement les structures actuellement disponibles et celles à venir apportera davantage d'informations sur l'adaptation halophile. En ce qui concerne l'étude de la structure des molécules d'eau à la surface de CaMalDH, comme la taille des cristaux de CaMalDH le permet, je pense que mener une expérience de diffraction neutronique pourrait permettre une compréhension plus fine de la structuration de l'eau en réseau de polygones.

doi:10.1016/j.jmb.2010.09.055 J. Mol. Biol. (2010) 404, 493–505

Contents lists available at www.sciencedirect.com

Journal of Molecular Biology

journal homepage: http://ees.elsevier.com.jmb

Gradual Adaptive Changes of a Protein Facing High Salt Concentrations

Nicolas Coquelle[1], Romain Talon[1], Douglas H. Juers[1,2], Éric Girard[1], Richard Kahn[1] and Dominique Madern[1]*

[1]*IBS, Institut de Biologie Structurale Jean-Pierre Ébel, Extremophilic and Large Molecular Assemblies Team, UMR 5075, CEA, CNRS, Université Joseph Fourier, 41 rue Jules Horowitz, F-38027 Grenoble, France*
[2]*Department of Physics, Whitman College, Walla Walla, WA 99362, USA*

Received 9 July 2010;
received in revised form
22 September 2010;
accepted 24 September 2010
Available online
1 October 2010

Edited by G. Schulz

Keywords:
halophilic;
malate dehydrogenase;
protein adaptation;
stability;
solubility

Several experimental techniques were applied to unravel fine molecular details of protein adaptation to high salinity. We compared four homologous enzymes, which suggested a new halo-adaptive state in the process of molecular adaptation to high-salt conditions. Together with comparative functional studies, the structure of malate dehydrogenase from the eubacterium *Salinibacter ruber* shows that the enzyme shares characteristics of a halo-adapted archaea-bacterial enzyme and of non-halo-adapted enzymes from other eubacterial species. The *S. ruber* enzyme is active at the high physiological concentrations of KCl but, unlike typical halo-adapted enzymes, remains folded and active at low salt concentrations. Structural aspects of the protein, including acidic residues at the surface, solvent-exposed hydrophobic surface, and buried hydrophobic surface, place it between the typical halo-adapted and non-halo-adapted proteins. The enzyme lacks inter-subunit ion-binding sites often seen in halo-adapted enzymes. These observations permit us to suggest an evolutionary pathway that is highlighted by subtle trade-offs to achieve an optimal compromise among solubility, stability, and catalytic activity.

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Halophilic microorganisms inhabit extremely saline environments, up to NaCl saturation. These microbes have selected various strategies of adaptation in order to cope with the resulting strong osmotic pressure.[1] One strategy is to maintain a low cytoplasmic salt concentration using salt-exclusion mechanisms together with the synthesis and/or accumulation of compatible solutes (osmolytes).[1] Halophilic organisms that use this strategy (typically eubacteria and higher organisms) have efficient transmembrane ion-selective pumps as well as specific pathways for osmolyte synthesis. A second strategy involves KCl accumulation at molar concentrations in the cytoplasm to counterbalance the external osmotic pressure. This strategy is typical of extreme halophilic archaea but was also recently discovered for the first time in a eubacterial species, *Salinibacter ruber* (*Sr*), which inhabits the same natural environment as halophilic archaea.[2] Whereas osmolytes typically have only small effects on enzyme stability and catalysis, high salt concentrations are generally deleterious to enzyme function.[3] Nevertheless, enzymes from halophilic cells have evolved to maintain solubility, catalytic efficiency, and structural stability in high concentrations of KCl, all necessary for proper function.[4,5]

*Corresponding author. E-mail address: Dominique.Madern@ibs.fr.

Abbreviations used: OAA, oxaloacetate; MalDH, malate dehydrogenase; LDH, lactate dehydrogenase; *Ca*, *Chloroflexus aurantiacus*; *Sr* (halophilic bacterium), *Salinibacter ruber* (halophilic bacterium); *Hm*, *Haloarcula marismortui* (halophilic archaea); *Tt*, *Thermus thermophilus*; AUC, analytical ultracentrifugation; PDB, Protein Data Bank.

0022-2836/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.

Work on several protein systems over the last 15 years has uncovered three general features of molecular adaptation via the KCl accumulation strategy.[6–10] First, halophilic proteins from extreme archaea are very acidic, with a mean calculated p*I* value of 5 for proteomes of various halophilic species. Crystal structures of several proteins from halophilic archaeal cells with KCl-rich cytoplasms have shown their surfaces to be rich in acidic residues.[11] Second, many halophilic proteins have specific ion-binding sites, often at subunit interfaces. Third, there is often a reduction of hydrophobic content of the protein, on the folded surface as well as buried in the core. This occurs, in some cases, via the substitution of large hydrophobic residues for small ones and, in other cases, concomitantly with the reduction of the number of lysine residues. A recent site-directed mutagenesis study underscored the decrease of a protein's solvent-accessible area as an important mechanism for halo-adaptation.[12]

One of the most extensively studied groups of proteins with respect to halo-adaptation is the family of lactate dehydrogenase (LDH)-like malate dehydrogenases (MalDHs).[13] A halophilic version of MalDH from the archaeon *Haloarcula marismortui* (*Hm*) is a 130-kDa homotetramer, bearing each of the three adaptive signatures described above.[14,15] The solvent interactions, stability, and weak protein–protein interactions of *Hm* MalDH in the presence of various salts have been studied in detail previously.[16] At high physiological salt concentrations, the solvation shell of this very acidic enzyme has a slightly higher concentration of salt than that of a non-halophilic protein, whereas its water binding capacity is comparable to that of a non-halophilic protein.[16,17] With a solvation shell similar to bulk solvent, the protein can remain soluble in high salt concentrations. At low salt concentrations (below 2 M KCl), the *Hm* MalDH enzyme dissociates and unfolds.[4,5] The stability of *Hm* MalDH depends on the types and concentrations of cations and anions present in the buffer.[15] Anions appear to control stability via a few strong binding sites, whereas cations play a dual role through both weak binding sites and hydration effects.

The crystal structure of *Hm* MalDH shows the surface of the tetramer to be enriched in acidic residues.[7,18] Complex salt bridges occur across the monomeric interfaces that form the tetramer, and each tetramer contains two sodium-binding sites and eight chloride-binding sites. Non-halophilic orthologs of LDH-like MalDH from several eubacteria have also been studied,[19] but to date, there is little information to provide a link between these enzymes and the halophilic *Hm* MalDH. To assess if an identical evolutionary constraint (high cytoplasmic KCl concentration) induced an adaptive response similar to that observed in *Hm* MalDH, we analyzed the molecular properties of LDH-like MalDH from the halophilic bacterium *Sr*.

Purification of this enzyme was described previously, and some of its biochemical properties were investigated. However, this characterization was only partial as the sequence and three-dimensional structure remained unknown at that time.[20] In the present study, we report the crystal structure of the enzyme at a resolution of 1.55 Å and analyze the solution behavior in different salt conditions. In addition, we extend and enhance this characterization by including two other non-halophilic counterparts. Our comparative work indicates that the *Sr* MalDH corresponds to a newly described acclimation state of an enzyme adapting to high-salt conditions.

Results

Comparison strategy

We explored the effects of KCl on (i) the conformational stability, (ii) weak protein–protein interactions that control solubility, and (iii) enzymatic activities of the various halophilic and non-halophilic homologs from the LDH and LDH–MalDH superfamily. We analyzed the data with respect to the available structural information including consideration of the structural features previously described as important for halophilic adaptation (Fig. 1). The halophilic LDH-like MalDHs were from the bacterium *Sr* and the archaeon *Hm*, while the non-halophilic enzymes were the bacteria *Chloroflexus aurantiacus* (*Ca*) LDH-like MalDH and *Thermus thermophilus* (*Tt*) LDH.[21]

Crystal structure of *Sr* MalDH

The structure of *Sr* MalDH enzyme was determined at 1.55 Å resolution using molecular replacement (Table 1). The asymmetric unit contains a 314-residue monomer, with the physiological tetramer obeying the crystal 222 symmetry (Fig. 1). The final *Sr* MalDH model shows a fold previously observed for enzymes of the LDH–MalDH family, composed of two topological domains: a dinucleotide-binding domain and a catalytic domain.[7,18,19] Sequence identities between *Sr* MalDH and *Tt* LDH, between *Sr* MalDH and *Ca* MalDH, and between *Sr* MalDH and *Hm* MalDH are 35%, 53%, and 34%, respectively, while structural similarities, as measured by RMSD, are 1.21 , 0.60 , and 1.07 Å, respectively, for 210, 258, and 216 superimposed C^{α} atoms. Weak electron density occurs in two short regions of the *Sr* enzyme. Consequently, the model does not include the two carboxy-terminal residues or, as frequently observed in crystal structures of LDHs and MalDHs,

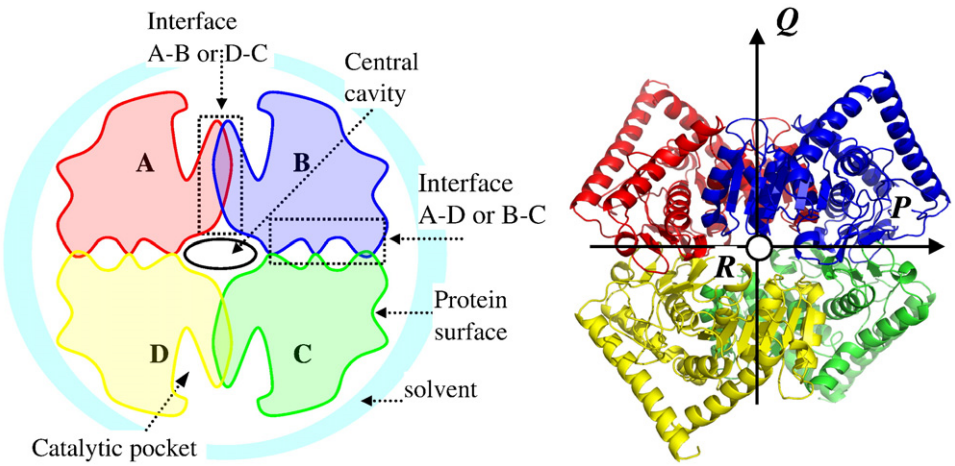


Fig. 1. Left panel: Schematic drawing of the tetrameric state of LDH and LDH-like MalDH. Each monomer is labeled A, B, C, or D. In the present study, several biochemical properties of the halophilic *Sr* LDH-like MalDH are analyzed with respect to several structural features: the nature of the various interfaces (A–B or B–C) that play a role in the quaternary structure, the protein surface, and the catalytic pocket. Right panel: C^{α} ribbon drawing of the tetrameric *Sr* enzyme. The tetramer has three dyads named P, Q, and R following the convention of Rossmann.

a flexible loop (residues 101–108 for *Sr* MalDH). The final *Sr* MalDH model also includes 340 water molecules per monomer with no ions.

Sr MalDH shows the conformational stability of a non-halophilic protein

The conformational stabilities of the four enzymes after a 24-h incubation at 25 °C in various KCl concentrations were compared. Data for *Tt* LDH and *Ca* MalDH were measured here, whereas data from *Sr* MalDH and *Hm* MalDH were taken from previous studies.[18,20] Both *Tt* LDH and *Ca* MalDH show a salt-independent stability, as expected for non-halophilic proteins (Fig. 2a and b). As was shown previously,[20] *Sr* MalDH is also stable over the whole KCl range (Fig. 2c), whereas the halophilic *Hm* MalDH unfolds below 2 M KCl (Fig. 2d). Previous measurements of the *Sr* MalDH quaternary assembly using analytical ultracentrifugation (AUC) showed that the enzyme is still tetrameric after 24 h of incubation at low salt,[4,6] which was repeated here. After the first AUC data acquisition on *Tt* LDH, *Ca* MalDH, and *Sr* MalDH at low KCl (0.05 M), the cell contents were homogenized and a second measurement was performed. The AUC results were identical before and after incubation, showing again that these enzymes are stable (data not shown).

The low-salt unfolding for *Hm* MalDH, which is well documented,[4,6] is due to the presence of ion-binding sites at the tetramer interfaces. A putative sodium-binding site at a dimer interface (two sites per tetramer) and two different chloride-binding sites (eight sites per tetramer) contribute to the native assembly.[7,18] The role of the chloride-binding sites in the stabilization and association of the enzyme has been well established using site-directed mutagenesis.[18,22]

Table 1. Crystal parameters, data collection, and refinement statistics

	Sr MalDH
Data collection	
ESRF beam line	BM30–ID29
Wavelength (Å)	0.9797–0.9320
Resolution range (Å)	33–1.55 (1.60–1.55)
Space group	*I*222
Cell dimensions (Å)	
a	76.11
b	87.72
c	100.45
R_{merge}	0.083 (0.499)
$\langle I/\sigma I \rangle$	11.4 (2.2)
Redundancy	6.4 (4.0)
Completeness (%)	97.5 (93.2)
Refinement	
Resolution (Å)	33–1.55
R_{work}/R_{free}	0.168/0.192
Number of atoms	
Protein	2360
Waters	340
Mean *B*-factor (Å^2)	
Protein	28.9
Water	41.7
RMSD	
Bond length (Å)	0.013
Bond angle (°)	1.33
Ramachandran statistics (%)	
Favorable	98.7
Allowed	1.3

Values in parentheses correspond to the highest-resolution shell.

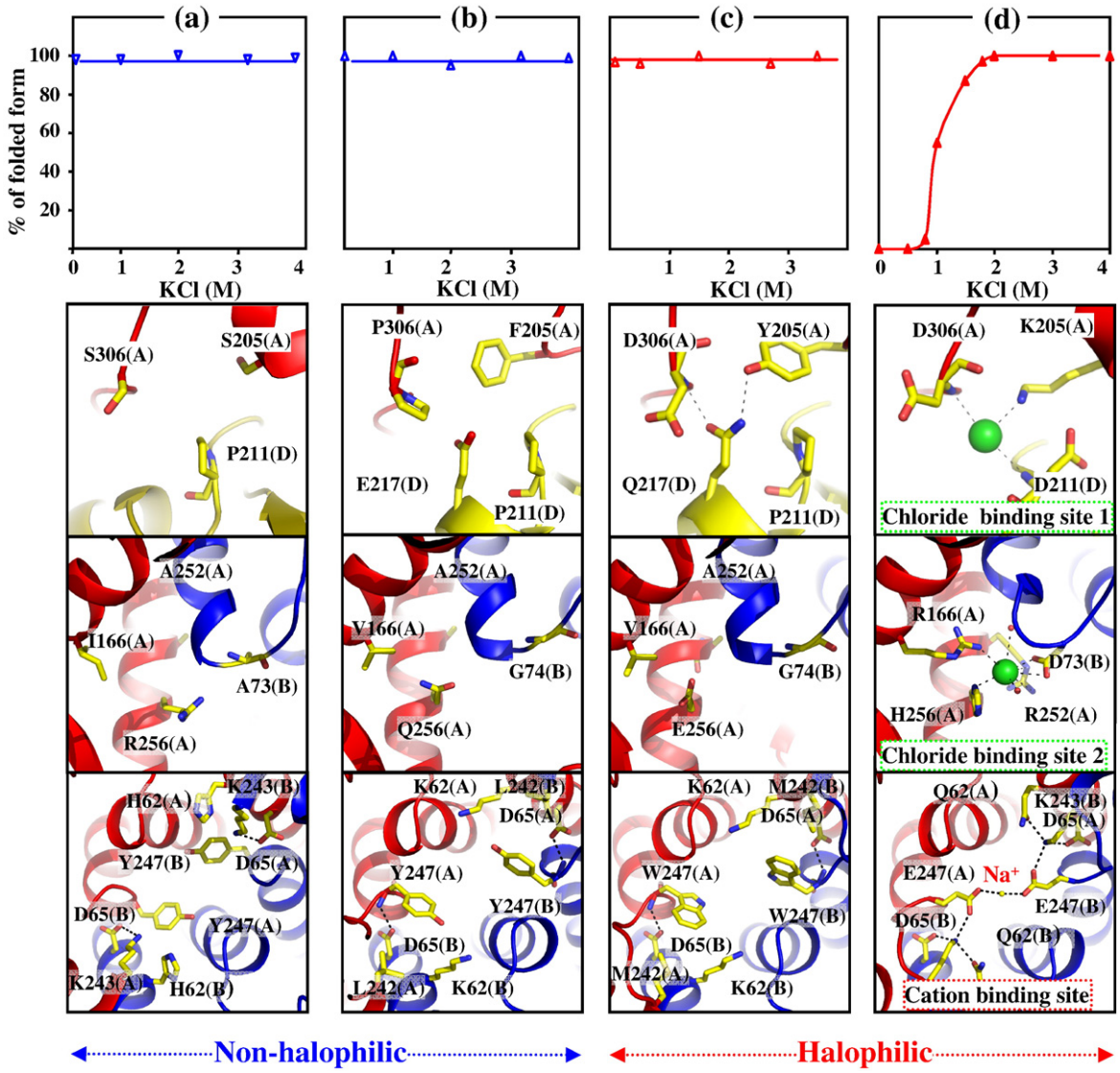


Fig. 2. Graphs: The effect of salt concentration on the conformational stability of *Tt* LDH (a), *Ca* (b), *Sr* (c), and *Hm* MalDH (d). Pictures: Close-up view of structures at positions that allow comparison of the most important amino acid substitutions involved in the salt-concentration-dependent stability. Non-halophilic enzymes (*Tt* LDH and *Ca* MalDH) and halophilic *Sr* MalDH lack the specific ion-binding sites that control the salt-concentration-dependent stability of the *Hm* enzyme: compare positions equivalent to those of *Hm* chloride-binding site no. 1 (four sites per tetramer) (top row), *Hm* chloride-binding site no. 2 (four sites per tetramer) (middle row), and *Hm* cation-binding site (two sites per tetramer) (bottom row). The monomers are colored according to Fig. 1.

At the positions equivalent to two different intersubunit chloride-binding sites (CB1 and CB2) of *Hm* MalDH (between monomers A and D), *Sr* MalDH does not contain ions and is more similar to its non-halophilic orthologs from *Tt* and *Ca*, which also lack bound ions (Fig. 2). According to structural criteria (ion–ligand distances and opposing negative charges), a sodium ion was identified in *Hm* MalDH between monomers A and B.[7,18] The same location in *Sr* MalDH is occupied by hydrophobic amino acids, suggesting that the cation-binding site is absent (Fig. 2). As lithium chloride, a non-physiological salt, was a constituent of the crystallization solutions used with *Sr* MalDH, a careful inspection of the solvent structure was conducted to detect these ions. Non-protein positive peaks (larger than 3 σ) in difference Fourier maps were modeled as water molecules. Due to the larger number of electrons for Cl^- (17) *versus* water (10), large positive peaks would be expected in the difference Fourier maps if chloride ions were bound to the protein. Moreover, a water molecule modeled in place of a chloride ion should have a very low *B*-factor compared to those of neighboring residues. No

chloride ions were identified in our model by using these criteria. Lithium ion–protein interactions can be characterized by interatomic distances and coordination geometries.[23] Based on high-resolution protein structures (<1.5 Å resolution), the distances between a bound lithium ion and its oxygen or nitrogen atom ligands are generally in the 1.8- to 2. 0-Å range. None of the water molecules bound to *Sr* MalDH showed such interaction distances, suggesting the absence of bound lithium ions. To distinguish water molecules from sodium ions (which differ by only a single electron), we considered the coordination geometry of solvent molecules with coordination number greater than 4, as well as structural similarities or differences between *Sr* and *Ca* MalDHs. No water molecule fulfilled the criteria to be assigned as a sodium ion. Finally, we tried to look for sites with unfavorable interactions such as negatively charged residues facing each other, as was used to identify a potential cation-binding site in *Hm* MalDH.[7,18] However, no such structural motif was observed in *Sr* MalDH. The other interactions at the various monomeric interfaces that build up the tetramer are very similar between *Sr* and *Ca* MalDH. The lack of both bound ions and unfavorable interactions explain why the conformational stability of *Sr* MalDH at low salt concentrations is similar to that of a non-halophilic ortholog.

Changes in surface properties control solubility

Effects of salt on the solubility of proteins are well studied.[24] For a protein to be soluble, the concentra-

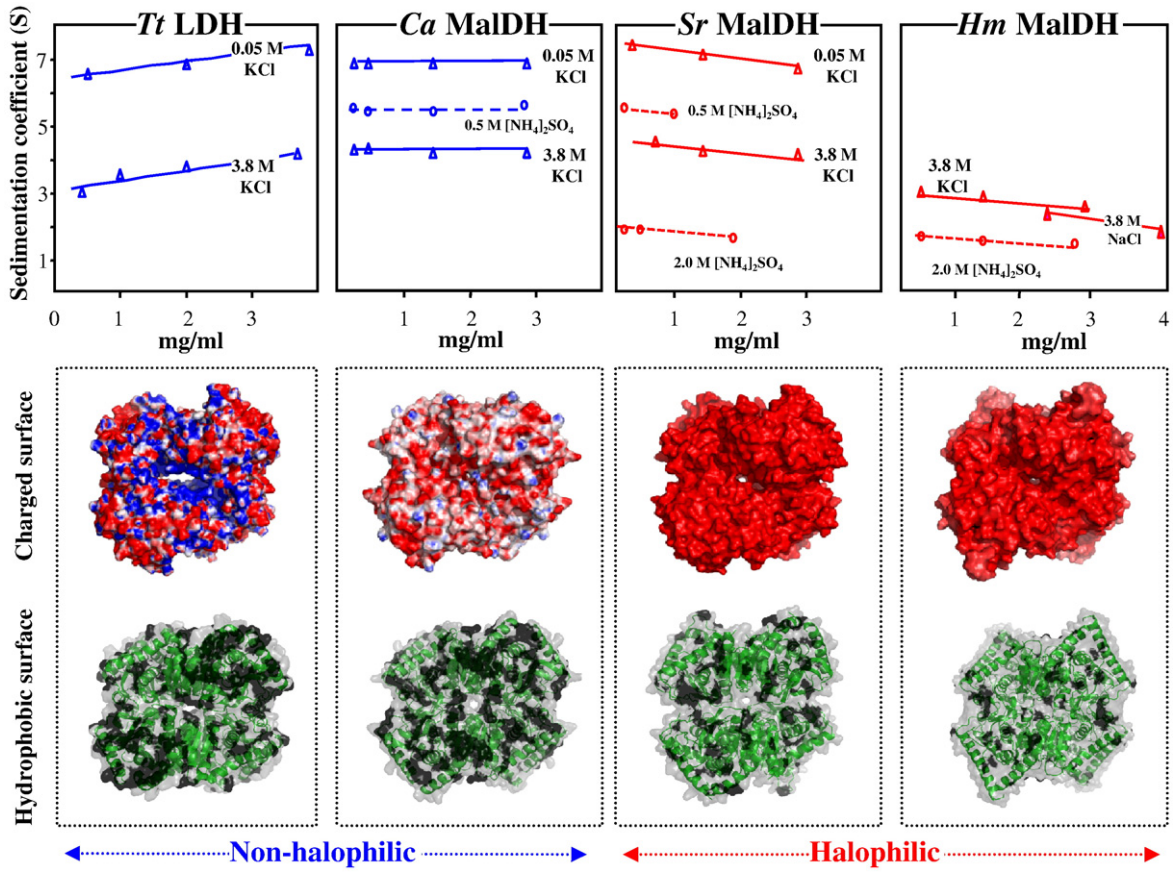


Fig. 3. Graphs: Variation of the experimental sedimentation coefficient (S_{exp}) at various salt concentrations demonstrating the repulsive nature of protein–protein contacts for the halophilic proteins (*Sr* MalDH and *Hm* MalDH) but attractive (or neutral) interactions for the non-halophilic protein (*Tt* LDH and *Ca* MalDH). The slopes were obtained from a linear regression of the S_{exp} data. Pictures: Surface representations of the tetrameric LDH and MalDH enzymes. Top: Electrostatic surface. The electrostatic potential was calculated at 300 K and 0 M salt concentration. Red corresponds to an electrostatic potential smaller than −517 mV. Blue corresponds to a potential greater than 517 mV. The calculated charge densities are −0.004, −0.006, −0.023, and −0.038 e Å^{-2} for *Tt* LDH, *Ca* MalDH, *Sr* MalDH, and *Hm* MalDH, respectively. Bottom: Hydrophobic surface representation. The solvent-accessible hydrophobic surface is colored black. The calculated values are 16,285, 13,811, 9487, and 7384 Å^2 for *Tt* LDH, *Ca* MalDH, *Sr* MalDH, and *Hm* MalDH, respectively. The ribbon drawing (green) is shown by transparency.

tion of salt in the solvation shell surrounding the protein should be similar to the bulk solvent. High salt strongly modifies the equilibrium between the bulk and the solvation shell, promoting either precipitation or dissolution when either a salting-out or a salting-in salt is added. For a given salt concentration, the tendency of a protein to precipitate (or to stay soluble) is determined by the nature of attractive (or repulsive) weak protein–protein interactions. AUC can be used to determine the dominant regime, because the sedimentation profiles are very sensitive to such interactions.[25] This method has been used previously with success in the study of halophilic proteins.[26] Under high-salt conditions, in which the enzyme is very soluble, the sedimentation coefficient decreases with increasing protein concentration, indicating a repulsive regime.[26] On the other hand, under salt conditions similar to those that promote crystallization of *Hm* MalDH, the sedimentation coefficient increases with protein concentration, indicating that interparticle interactions are under an attractive regime.

We applied this method to study the behavior of the MalDH and LDH enzymes in various KCl concentrations. The experimental sedimentation coefficients were determined using the software SEDFIT[27] (Materials and Methods). To avoid superimposition, we plotted them without correction for density and viscosity, which has no effect on the slope (Fig. 3). The non-halophilic and halophilic enzymes showed opposite behavior (Fig. 3). The non-halophilic enzyme from *Tt* is controlled by a strongly attractive regime at both low and high KCl concentrations (Fig. 3a), and at 3.8 M KCl, the enzyme starts to precipitate at 6–8 mg/ml. The *Ca* MalDH exhibits a more neutral behavior, with small or slightly negative slopes (Fig. 3b). At low salt concentrations, we were able to concentrate *Ca* MalDH to 20–30 mg/ml. In contrast, *Sr* MalDH exhibits a repulsive behavior (Fig. 3c), and at 3.8 M KCl, the protein remains soluble at 35 mg/ml. Due to the limited amount of protein available, we were not able to determine the solubility limit of this enzyme. *Hm* MalDH also exhibits repulsive behavior at high KCl concentrations, and at 4 M NaCl, this enzyme can be concentrated to 200 mg/ml.[17] For probing the relationship between various enzyme surface properties and solubility in more drastic salt conditions, we also measured the AUC profiles in $(NH_4)_2SO_4$ (0.5 and 2.0 M), which typically strongly favors protein precipitation. Halophilic *Sr* MalDH and *Hm* MalDH are still under a repulsive regime and stay soluble (Fig. 3c and d). With *Tt* MalDH, data are very difficult to record at 0.5 and 2.0 M $(NH_4)_2SO_4$, because the *Tt* enzyme precipitates (slowly at 0.5 M and immediately at 2 M). In the case of *Ca* MalDH, the slope indicates a slightly attractive or neutral behavior at low salt concentrations (Fig. 3b). At 2.0 M, $(NH_4)_2SO_4$ data cannot be recorded due to the strong precipitation of the protein.

In a given solvent, differences in protein–protein and protein–solvent interactions between homologous enzymes are mainly due to the nature of the protein surfaces. We therefore analyzed the surfaces of *Tt* LDH and *Ca*, *Sr*, and *Hm* MalDH to determine the structural consequences of the acidic amino acid substitutions in halophilic enzymes. The *Sr* enzyme has about the same number of acidic amino acids oriented toward the solvent (27) as *Hm* (25) and has 14 and 18 more than the *Ca* and *Tt* enzyme, respectively. Of the exposed acidic residues in *Sr* and *Hm* MalDH, 12 are located in identical positions. The qualitative difference between non-halophilic and halophilic enzymes is reflected in the surface electrostatic potential, which is more negative for the halophilic enzymes (Fig. 3, middle row). For the non-halophilic enzymes, the calculated charge density is -0.004 e Å^{-2} for *Tt* MalDH and -0.006 e Å^{-2} for *Ca* MalDH, while it is -0.023 and -0.038 e Å^{-2} for halophilic *Sr* and *Hm* MalDH, respectively.

The solubility should also be impacted by the hydrophobic surface, since the hydrophobic effect is enhanced by high concentrations of neutral salts or by salting-out salts.[3] The total surface area of the tetramer is about the same (within 2% of 41,000 Å^2) for each enzyme, but the exposed hydrophobic surface decreases from 38% of the total area in the case of the *Tt* LDH to 33%, 23%, and 18% for the *Ca*, *Sr*, and *Hm* enzymes, respectively (Fig. 3, bottom panels). Our data thus show a strong correlation between the nature of intermolecular interactions (attractive *versus* repulsive) and surface physical properties (electrostatic charge and hydrophobicity).

This suggests that the main adaptive consequence of acidic amino acid substitutions in a halophilic protein favors a change in its intrinsic solubility due to the modification of its surface properties.

The favorable change in enzymatic activity at high salt concentration correlates with subtle structural changes

We measured the effect of KCl on the activity of each enzyme. As expected for non-halophilic enzymes, the activities of *Tt* LDH and *Ca* MalDH were drastically inhibited at KCl concentrations higher than 2 M (Fig. 4a and b). For *Hm* MalDH (Fig. 4d), the activity was maximal over the salt concentration range in which the enzyme is stable (2–4 M). At lower KCl, activity cannot be recorded because the enzyme deactivates.[4] The *Sr* MalDH activity shows an intermediate behavior, as it drops with increasing KCl but still remains at 60% of maximum at high KCl (Fig. 4c).

Structural analysis showed that some structural changes correlate with the increase of activity at

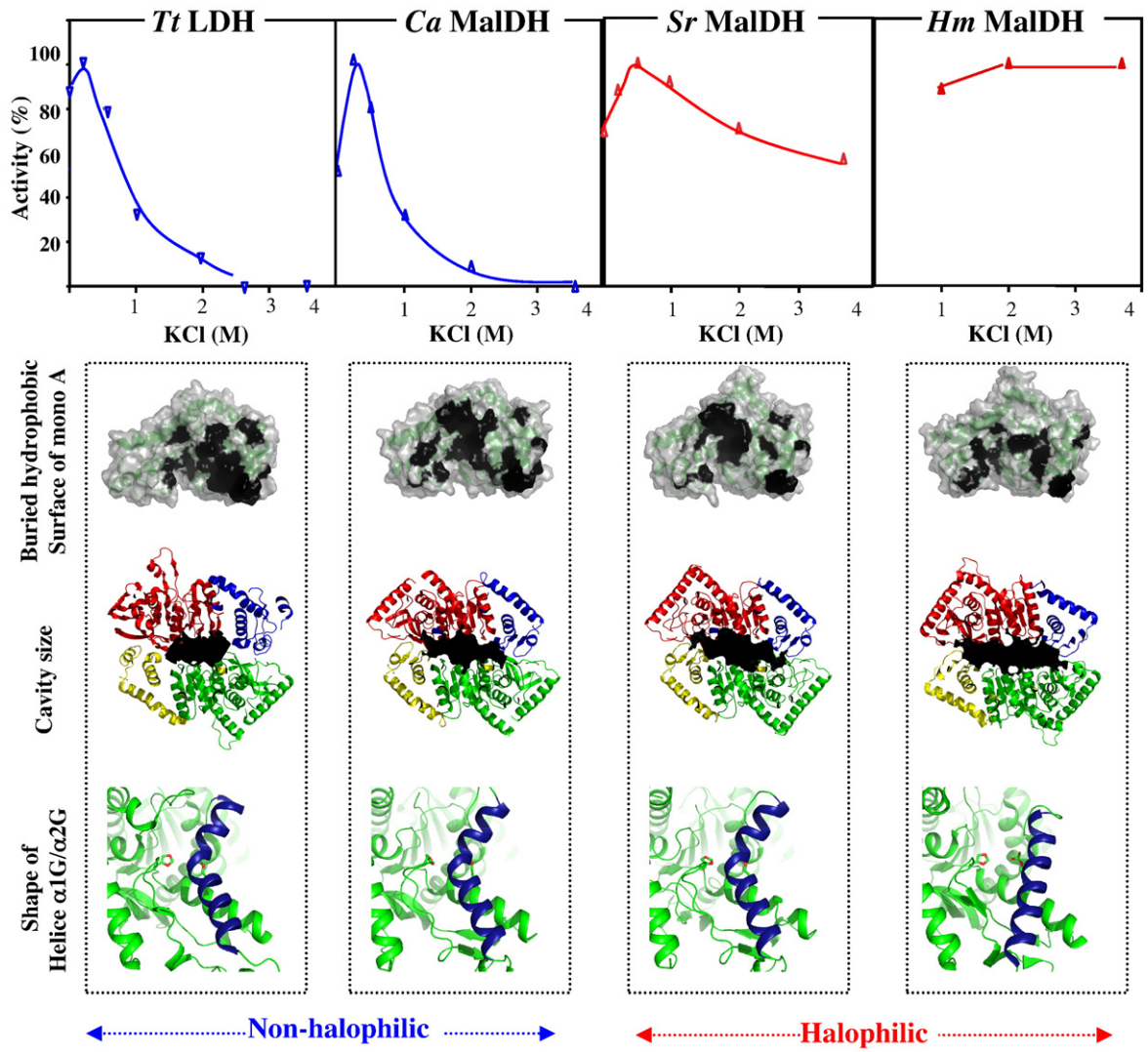


Fig. 4. Graphs: The effect of KCl concentration on the enzymatic activity of *Tt* LDH, *Ca* MalDH, *Sr* MalDH, and *Hm* MalDH. Pictures: Illustrations of structural differences related to activity. Top: Surface contact of monomer A at the monomer–monomer interface that build up the dimer [A–B] or [D–C] (see also Fig. 1). The hydrophobic buried surface area is colored black. The percentage of the hydrophobic surface that contributes to the total buried surface is 60%, 56%, 47%, and 38% for *Tt* LDH, *Ca* MalDH, *Sr* MalDH, and *Hm* MalDH, respectively. Second row: Central cavity (in black) of the enzymes. Calculated volume is 1775, 1602, 2156, and 3516 Å^3 for *Tt* LDH, *Ca* MalDH, *Sr* MalDH, and *Hm* MalDH, respectively. Third row: Close-up view of the catalytic pocket with residues (His195, Arg171, red sticks) involved in catalysis; the backbone is colored green.

high salt concentrations. According to the folding and association pathway of LDH and LDH-like MalDH, the minimal folded unit that sustains activity assembles via the association of monomer A (or D) with monomer B (or C) to build the dimer [A–B] or [D–C].[28] The total surface area of each dimer, as well as the area buried at the dimer interface, is about the same (±5%, not shown). However, the composition of the monomer–monomer interface changes with the enzyme. Specifically, the hydrophobic buried surface represents 60% and 56% of the total buried surface of this interface in the non-halophilic *Tt* LDH and *Ca* MalDH enzymes and decreases to 47% and 38% for *Sr* and *Hm* MalDH, respectively (Fig. 4, second row from top). Assembly of the active dimers that build up the tetramer creates a solvent-accessible cavity (Fig. 4, third row from top). The volume of this cavity increases from an average of 1688 Å^3 for the non-halophilic enzymes to 2156 and 3513 Å^3 for *Sr* and *Hm* MalDH, respectively.

Inside the catalytic vacuole, there are only a few relatively conservative substitutions that are not involved in substrate binding and reactivity and

are therefore probably not responsible for the changes in catalytic efficiency. The catalytic mechanism of LDH and MalDH is well understood, following an ordered mechanism in which NADH binds first inside the catalytic vacuole followed by the substrate.[29] In LDH, substrate binding and release induce structural reorganizations of different amplitudes between the apo-state and the ternary complex.[30,31] When switching from the apo-state to the ternary complex, there is a large movement of the catalytic loop and helix αD/E associated with other reorganizations of less importance involving (i) the connecting region between strand βJ and helix α1G and (ii) helix α1G/α2G. In the non-halophilic *Tt* LDH, the connecting region corresponds to a small helix αT, which is shortened in the non-halophilic *Ca* MalDH. In the halophilic *Sr* MalDH, this very small helix superimposes very well with the one observed in *Ca* MalDH (not shown). In *Hm* MalDH, the connecting region βJ–α1G does not contain any secondary elements due to the deletion of few residues. This has a consequence on the location and shape of the α1G/α2G helix, which is in a straight position, allowing the catalytic site to be in a wider open state, a situation that favors substrate trafficking.[32] This phenomenon is not observed in the halophilic *Sr* MalDH enzyme. As observed with its non-halophilic counterparts, the α1G/α2G helix of *Sr* MalDH is kinked (Fig. 4, bottom row). Our data suggest that gradual structural rearrangements promote favorable change of enzymatic activity at high salt concentrations.

Flexibility of the *Sr* MalDH catalytic pocket

One way that the *Sr* enzyme could acquire increased activity at high salt concentrations relative to the *Ca* enzyme would be to increase overall flexibility in the active site region. Because of their effects on local conformational entropy, specific glycine and proline substitutions have been emphasized in adaptive mechanisms within the malate/LDH superfamily.[21,33,34] In cold-adapted LDHs, glycine residues have been selected for mostly in hinge regions that control enzyme conformational flexibility.[21] We found substitutions of non-glycine residue from *Ca* MalDH into glycine residues in *Sr* MalDH. These changes occur mainly in regions surrounding the catalytic pocket that undergo large conformational changes associated with catalysis (Fig. 5a). The conformational entropy is further increased in the *Sr* enzyme relative to the *Ca* one via the removal of Pro at positions 107 and 196 (Fig. 5a).

We also performed a comparison of the refined crystallographic *B*-factors between *Sr* and *Ca* MalDH, using the relative *B*-factor difference at equivalent positions, which limits bias coming from data collection, data integration, and refinement processes. This approach was successful in highlighting the role of three loops in the catalytic process of urate oxidase.[35] Changes in *B*-factors at equivalent

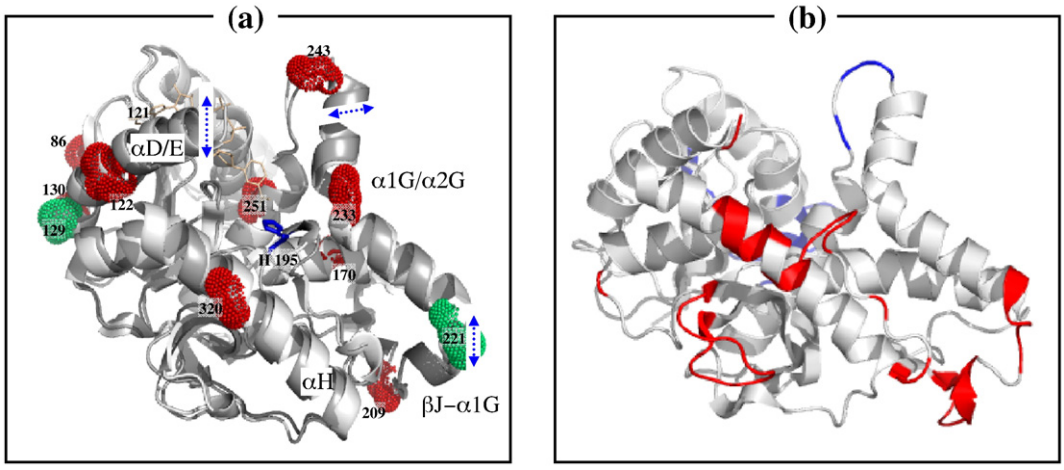


Fig. 5. (a) Ribbon diagram superimposition of *Sr* (gray) and *Ca* MalDH (pale gray). In the *Ca* MalDH structure, the NADH is colored yellow and the catalytic residue H195 is in blue. The blue arrows indicate the main structural differences between the apo-state and the ternary complex of LDH. The locations of additional glycine residues (red balls) in *Sr* MalDH compared to *Ca* MalDH and locations of proline residue (green) in *Ca* MalDH compared to *Sr* MalDH are indicated (PDB numbering). (b) View of the *Ca* MalDH backbone showing in color those residues whose relative crystallographic *B*-factor difference between the *Ca* MalDH and *Sr* MalDH structures is greater than 1 standard deviation with respect to the average value. Red indicates that the *Sr* structure has the higher relative *B*-factor, while blue indicates that the *Ca* structure has the higher relative *B*-factor.

positions between the two enzymes convey the effects of amino acid substitutions on protein flexibility, including glycine and proline substitutions as was discussed previously. According to our calculations, there is a net increase of flexibility in *Sr* MalDH compared to *Ca* MalDH (Fig. 5, bottom panel). It should be pointed out that flexibilities of various regions surrounding the catalytic vacuole of *Sr* MalDH are increased, in particular the connecting region βJ–α1G. In the case of LDH, it has been demonstrated that increasing the glycine content in this region permits the enzyme to be more efficient.[31,32]

Discussion

The intermediate nature of the structure and function of *Sr* MalDH offers new insight into molecular halophilic adaptation. The comparison helps to elucidate the specific roles of different types of amino acid substitutions between halophilic and non-halophilic enzymes. The high KCl concentration induces two types of nonspecific effects on enzyme structure and function that are relevant here. First, weak protein–protein interactions tend to become more attractive due to preferential hydration, decreasing protein solubility.[14,16] Second, protein stability is increased not only because of preferential hydration but also because the hydrophobic effect is enhanced in high salt.[36] The increase in stability is thought to freeze out conformational fluctuations required for catalytic activity.[4,5] Thus, the function of a protein placed in a halophilic environment can be improved by (a) increasing solubility and (b) increasing flexibility via decreasing stability, which restores catalytic activity.

With *Sr* and *Hm* MalDH, the solubility problem for the native tetrameric state is solved in two ways. First, the surface is enriched in acidic amino acids, which is well described. About half of the acidic residues occur at the same surface positions, suggesting that either both enzymes had a common halophilic protein ancestor with an intermediate acidic surface (divergent evolution) or there are some specific key surface positions for the location of charge to enhance solubility (convergent evolution). Our comparison and previous studies[14,16,24] show that the acidic enrichment in halophilic proteins enhances both weak (repulsive) interparticle interactions and protein–solvent interactions. Selection of an acidic surface is therefore needed to maintain protein solubility in high salt. Second, by reducing the external hydrophobic surface of the folded protein, hydrophobic protein–protein attraction is also diminished. The reduction in positively charged residues (mainly lysine) sometimes observed in halophilic proteins thus improves solubility by both increasing the net negative surface charge and decreasing the hydrophobic surface area of the folded protein.

In the LDH–MalDH superfamily, solving the activity problem at high salt concentrations is consistent with the notion of stability/activity trade-offs, a phenomenon previously observed in the case of thermal adaptation. Numerous studies suggest that enzymes adapted to high temperatures have reduced activity at lower temperatures because of reduced flexibility.[33,37–39] In terms of the stability/activity landscape, increasing the salt concentration for halo-adaptation corresponds to decreasing the temperature for cold adaptation. High salt concentrations as well as low temperatures decrease the rate at which the enzyme reaches the conformational subset favorable to sustain activity. In both cases, the effects of adaptive substitutions involve increase flexibility, restoring activity to the hyperstabilized enzyme. In the case of LDH-like MalDH, previous studies have shown that the minimal unit that sustains enzymatic activity corresponds to the dimer [A–B] or [D–C].[28,40] Here, our results suggest that the decrease of buried hydrophobic surface between monomers may favor a higher flexibility of the active dimeric unit for the halo-adapted enzymes. We point out, however, that because the native state of oligomeric proteins corresponds to the final step of a complex assembly process requiring the proper folding of intermediates and their association into oligomers, we cannot exclude the effect of the substitutions on this pathway. In particular, substitutions responsible for decreasing the hydrophobic surface between monomers that compose active LDH-like MalDH dimers could have been also selected to prevent improper association due to the high salt concentration. Another potential solution to the activity problem via increased flexibility is with glycine substitutions and decrease of proline content, which has been observed in cold-adapted LDH.[20,31] Although similar changes have occurred in *Sr* MalDH, the sum of these various effects (change of surface composition, decrease of buried hydrophobic interface, and increase of local entropy) is not enough to have fully optimized the enzymatic efficiency of *Sr* MalDH at high salt concentrations, since the activity is still inhibited by 50% compared to its maximal activity achieved at low salt concentration.

Our work suggests that the evolutionary process of proteins adapting to high salt concentrations is an "ordered/sequential" trade-off between activity, solubility, and stability to counterbalance the unfavorable effect of high salt concentrations on enzyme conformational landscape fluctuations. In the first step of halo-adaptation, sequence changes are selected to favor solubility and partially prevent salt inhibition. The activity at high salt concentrations can then be improved, further reducing stability and correspondingly increasing the flexibility required for catalysis.

Materials and Methods

Sample preparation

Purifications of *Sr*, *Ca*, and *Hm* MalDH and *Tt* LDH enzymes were carried out as previously reported.[6,19–21] Protein concentrations were determined from the absorbance at 280 nm using extinction coefficients calculated from the sequence of each protein†.

Conformational stability

Residual circular dichroism (CD) and residual activity were used to determine conformational stability.[28] The proteins of interest were incubated for 24 h at 0.1 mg/ml in various KCl concentrations prior to performing CD and activity assays.

Standard enzymatic assay

The activity was determined by following the oxidation of NADH over 30 s by measuring the decrease in absorbance at 340 nm (Beckman DU 7500 Spectrophotometer). For MalDH, the measurements were performed at 60 °C (near the optimal temperature for each enzyme) using saturating concentrations of the substrate oxaloacetate (OAA) to avoid artifacts from K_m variation due to high salt concentrations.[4] For the measurements on *Tt* LDH, the temperature was set to 75 °C and the substrate was pyruvate instead of OAA. Enzyme aliquots were mixed with 1 ml of 50 mM Tris–HCl, pH 8.0, supplemented with OAA (0.3 mM) or pyruvate (0.3 mM) and NADH (0.2 mM) at various salt concentrations.

Sedimentation velocity measurement

Sedimentation velocity experiments were performed with a Beckman XLI analytical ultracentrifuge, equipped with a UV scanning system, using an eight-hole AN-60 Ti rotor with a double-channel centerpiece of 1.2 cm optical path length. For each sample, 200 absorbance profiles were recorded at 280 nm (*Tt* LDH, *Sr* MalDH, and *Hm* MalDH) or 275 nm (*Ca* MalDH) at 42,000 RPM, at 20 °C after a temperature equilibration of 2 h. At a protein concentration close to 3 mg/ml, data were also recorded at 250 nm to avoid optical saturation artifacts. The experimental sedimentation coefficients were determined using the continuous size distributions $c(s)$ option in the SEDFIT software.[27] The data were fitted over the range of 0.1 to 15 s with a resolution grid of 200 points and systematic noise subtraction. The sedimentation profiles were well fit with a single component (>95% of the signal) and some discrete non-interacting species, which are considered to be contaminants. The experimental S value was obtained by the weight-average S value of the main sharp-distribution peak. Because the analyzed enzymes are homologous and display a globular shape, we assume that they have a same partial specific volume of 0.74 and friction coefficient f/f_0 of 1.20. In the case of *Hm* MalDH, this last assumption is only valid at high salt concentrations in which the enzyme is stable. According to SEDFIT specifications, under optimal conditions, the S values are given with errors of ±0.4%. We assume that, in most of our experiments, the small RMSD and low residual variation of ±0.015 optical density permitted us to determine S values with an uncertainty of ±1%. The $c(s)$ analysis within such confident limits was shown to be very satisfactory to characterize halophilic *Hm* MalDH in complex solvents[26] and to probe the behavior of an elongated protein in concentrated salt solution.[41]

Crystallization and structure determination of *Sr* MalDH

Crystals were grown at room temperature using the hanging-drop vapor-diffusion method. *Sr* MalDH protein (11 mg/ml in 50 mM Tris–HCl, pH 8) was mixed with an equal volume (2 µl) of mother liquor of 1 M LiCl, pH 4.0, 0.1 M citric acid, and 10% polyethylene glycol 6000. The crystal was soaked into a cryo-protective buffer [mother liquor containing 20% (v/v) added glycerol] before being flash-cooled in a nitrogen vapor stream at 100 K. Diffraction data were collected to a resolution of 1.55 Å at beam line BM30 of the European Synchrotron Radiation Facility (ESRF). Ice rings at resolutions of 2.2 and 1.9 Å were visible in diffraction frames and reflections in these resolution ranges were excluded. A second crystal was soaked in a slightly different cryo-protectant buffer [mother liquor containing 25% (v/v) added glycerol] and diffraction data were recorded to a resolution of 1.8 Å at ESRF beam line ID29. Diffraction data frames from each data set were indexed and integrated with the program XDS. As the crystals were isomorphous (a difference of 0.1%, 0.4%, and 0.3% in cell parameters was observed for a, b, and c, respectively), the two data sets were scaled and merged using the program XSCALE in space group $I222$ with cell dimensions of 76.11, 87.72, and 100.45 Å. Structure factor amplitudes were derived using the program XDSCONV[42] (Table 1). The asymmetric unit content was estimated as a single subunit via the calculation of the Matthews coefficient.[43]

The three-dimensional structure was determined using the molecular replacement method with the program PHASER[44] (v 1.3.3) and diffraction data in the resolution range 40–3 Å. The search model was a monomer of *Chlorobium tepidum* MalDH [Protein Data Bank (PDB) code 1GUZ] and a unique solution was found with a Z-factor of 26.3 (LL gain of 613). After a rigid-body refinement using data in the resolution 40–3 Å, an initial R-factor of 42% was obtained. Subsequent refinement process consisted of energy minimization and individual anisotropic B factor refinement for protein atoms and isotropic B-factor refinement for solvent atoms using the complete set of recorded reflections using the program PHENIX[45] (v 1.6_288). The model was built with Coot[46] between refinement cycles. The current refined atomic model comprises 306 residues and 340 water molecules. The homotetrameric molecule that corresponds to the biological assembly can be generated by the crystal symmetry operators of the $I222$ space group. The quality

† http://www.expasy.ch/tools/protparam.html

of the model was checked using the MolProbity server[47] (Table 1). The numbering of *Sr* MalDH model is based on that of *Hm* MalDH and the Eventoff nomenclature.[48]

Model analysis

The model of *Sr* MalDH was compared to the structures of the halophilic *Hm* MalDH (PDB code 1GT2) and of the non-halophilic *Ca* MalDH (PDB code 1GUY) and *Tt* LDH (PDB code 2V6M). Residue numbering of *Ca* MalDH in this article is based on *Sr* MalDH numbering and differs from that of the initial PDB. The solvent-accessible surface area of all tetrameric MalDHs was calculated using the program AREAIMOL.[49] For each tetramer, interfaces between monomers were examined with the program PISA.[50] The probe-accessible volume of the central cavities was calculated using the program VOIDOO.[51] The probe radius was set to 1.4 Å (i.e., water radius) and the central cavity was specifically selected with coordinates of a cavity point. The electrostatic potential was generated with the program APBS[52] and rendered with PyMOL (PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC), which was also used to prepare all the figures. The relative crystallographic *B*-factor difference for each residue was calculated for the backbone atoms using the formula $\frac{(B_{Ca} - \langle B_{Ca}\rangle) - (B_{Sr} - \langle B_{Sr}\rangle)}{\langle B_{Ca}\rangle - \langle B_{Sr}\rangle}$, where B_{Ca} and B_{Sr} refer to *B*-factor values extracted from *Ca* and *Sr* MalDH models, respectively, and where $\langle B\rangle$ refers to the mean value for the whole structure. This calculation leads to a distribution with a zero mean value and limits bias from refinements. The *Sr* MalDH sequence was aligned to the sequence from *Ca* MalDH.

Accession numbers

Coordinates and structure factors have been deposited in the PDB with accession number 3NEP.

Acknowledgements

We thank Dr Giuseppe Zaccai for helpful discussions and critical reading of the manuscript. Financial support by the Centre National de la Recherche Scientifique [Project EOPV (Environnements Planétaires et Origines de la Vie)] is acknowledged.

References

1. Oren, A. (2002). Halophilic microoganisms and their environments. In *Cellular Origin and Life in Extreme Environments* Kluwer Academic Publishers, Dordrecht, The Netherlands.
2. Antón, J., Oren, A., Benlloch, S., Rodríguez-Valera, F., Amann, R. & Rosselló-Mora, R. (2002). *Salinibacter ruber* gen. nov., sp. nov., a novel, extremely halophilic member of the *Bacteria* from saltern crystallizer ponds. *Int. J. Syst. Evol. Microbiol.* **52**, 485–491.
3. Timasheff, S. N. (1991). Solvent effect on protein stability. *Curr. Opin. Struct. Biol.* **2**, 35–39.
4. Madern, D., Ebel, C. & Zaccai, G. (2000). Halophilic adaptation of enzymes. *Extremophiles*, **4**, 91–98.
5. Mevarech, M., Frolow, F. & Gloss, L. M. (2000). Halophilic enzymes: proteins with a grain of salt. *Biophys. Chem.* **86**, 155–164.
6. Bonneté, F., Madern, D. & Zaccai, G. (1994). Stability against denaturation mechanisms in halophilic malate dehydrogenase "adapt" to solvent conditions. *J. Mol. Biol.* **244**, 436–447.
7. Richard, S. B., Madern, D., Garcin, E. & Zaccai, G. (2000). Halophilic adaptation: novel solvent protein interactions observed in the 2.9 and 2.6 Å resolution structures of the wild type and a mutant of malate dehydrogenase from *Haloarcula marismortui*. *Biochemistry*, **39**, 992–1000.
8. Bieger, B., Essen, L. O. & Oesterhelt, D. (2003). Crystal structure of halophilic dodecin. A novel dodecameric flavin binding protein from *Halobacterium salinarum*. *Structure*, **4**, 375–385.
9. Britton, K. L., Baker, P. J., Fisher, M., Ruzheinikov, S., Gilmour, D. J., Bonete, M. J. *et al.* (2006). Analysis of protein solvent interactions in glucose dehydrogenase from the extreme halophile *Haloferax mediterranei*. *Proc. Natl Acad. Sci. USA*, **103**, 4846–4851.
10. Paul, S., Bag, S. K., Das, S., Harvill, E. T. & Dutta, C. (2008). Molecular signature of hypersaline adaptation: insights from genome and proteome composition of halophilic prokaryotes. *Genome Biol.* **9**, R70.
11. Wende, A., Johansson, P., Vollrath, R., Dyall-Smith, M., Oesterhelt, D. & Grininger, M. (2010). Structural and biochemical characterization of a halophilic archaeal alkaline phosphatase. *J. Mol. Biol.* **400**, 52–62.
12. Tadeo, X., López-Méndez, B., Trigueros, T., Laín, A., Castaño, D. & Millet, O. (2009). Basis for the aminoacid composition of proteins from halophilic archea. *PLoS Biol.* **7**, e1000257.
13. Madern, D. (2002). Molecular evolution within the L-malate and L-lactate dehydrogenase super-family. *J. Mol. Evol.* **54**, 825–840.
14. Ebel, C., Costenaro, L., Pascu, M., Faou, P., Kernel, B., Proust-De Martin, F. & Zaccai, G. (2002). Solvent interactions of halophilic malate dehydrogenase. *Biochemistry*, **41**, 13234–13244.
15. Ebel, C., Faou, P., Kernel, B. & Zaccai, G. (1999). The relative role of anions and cations in the stabilization of halophilic malate dehydrogenase. *Biochemistry*, **38**, 9039–9047.
16. Costenaro, L., Zaccai, G. & Ebel, C. (2002). Link between protein–solvent and weak protein–protein interactions gives insight into halophilic adaptation. *Biochemistry*, **41**, 13245–13252.
17. Bonneté, F., Ebel, C., Zaccai, G. & Eisenberg, H. (1993). Biophysical study of halophilic malate dehydrogenase in solution: revised subunit structure and solvent interactions of native and recombinant enzyme. *J. Chem. Soc., Faraday Trans.* **89**, 2659–2666.
18. Irimia, A., Ebel, C., Madern, D., Richard, S. B., Cosenza, L. W., Zaccai, G. & Vellieux, F. M. (2003). The oligomeric state of *Haloarcula marismortui* malate dehydrogenase are modulated by solvent components as shown by crystallographic and biochemical studies. *J. Mol. Biol.* **326**, 859–873.

19. Dalhus, B., Saarinen, M., Sauer, U. H., Eklund, P., Johansson, K., Karlsson, A. *et al.* (2002). Structural basis for thermophilic protein stability: structures of thermophilic and mesophilic malate dehydrogenases. *J. Mol. Biol.* **318**, 707–721.
20. Madern, D. & Zaccai, G. (2004). Molecular adaptation: the malate dehydrogenase from the extreme halophilic bacterium *Salinibacter ruber* behaves like a non-halophilic protein. *Biochimie*, **86**, 295–303.
21. Coquelle, N., Fioravanti, E., Weik, M., Vellieux, F. & Madern, D. (2007). Activity, stability and structural studies of lactate dehydrogenases adapted to extreme thermal environments. *J. Mol. Biol.* **374**, 547–562.
22. Madern, D. & Ebel, C. (2007). Influence of an anion-binding site in the stabilization of halophilic malate dehydrogenase from *Haloarcula marismortui*. *Biochimie*, **89**, 981–987.
23. Olsher, U., Izatt, R. M., Bradshaw, J. S. & Dalley, N. K. (1991). Coordination chemistry of lithium ion, a crystal and molecular structure review. *Chem. Rev.* **91**, 137–164.
24. Tardieu, A, Bonneté, F., Finet, S. & Vivarès, D. (2002). Understanding salt or PEG induced attractive interactions to crystallize biological macromolecules. *Acta Crystallogr., Sect. D.: Biol. Crystallogr.* **58**, 1549–1553.
25. Behlke, J. & Ristau, O. (1999). Analysis of the thermodynamic non-ideality of proteins by sedimentation equilibrium experiments. *Biophys. Chem.* **76**, 13–23.
26. Solovyova, A., Schuck, P., Costenaro, L. & Ebel, C. (2001). Non-ideality by sedimentation velocity of halophilic malate dehydrogenase in complex solvents. *Biophys. J.* **81**, 1868–1880.
27. Schuck, P. (2000). Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys. J.* **78**, 1606–1619.
28. Madern, D., Ebel, C., Mevarech, M., Richard, S. B., Pfister, C. & Zaccai, G. (2000). Insights into the molecular relationships between malate and lactate dehydrogenases. Structural and biochemical properties of monomeric and dimeric intermediates of a mutant of tetrameric LDH-like malate dehydrogenase from the halophilic archaeon *Haloarcula marismortui*. *Biochemistry*, **39**, 1001–1010.
29. Holbrook, J. J., Lilias, A., Steindel, S. J. & Rossmann, M. G. (1975). Lactate dehydrogenase. In *The Enzymes* (Boyer, P. D., ed.), pp. 191–292, Academic Press, New York, NY.
30. Gerstein, M. & Chothia, C. (1991). Analysis of protein loop closure. Two types of hinges produce one motion in lactate dehydrogenase. *J. Mol. Biol.* **220**, 133–149.
31. Iwata, S., Kamata, K., Yoshida, S., Minowa, T. & Ohta, T. (1994). T and R states in the crystals of bacterial L-lactate dehydrogenase reveal the mechanism for allosteric control. *Nat. Struct. Biol.* **1**, 176–185.
32. Dym, O., Mevarech, M. & Sussman, J. L. (1995). Structural features that stabilize halophilic malate dehydrogenase from an archaebacterium. *Science*, **267**, 1344–1346.
33. Fields, P. A. & Somero, G. N. (1999). Hot spots in cold adaptation: localized increases in conformational flexibility in lactate dehydrogenase A4 orthologs of Antarctic notothenioid fishes. *Proc. Natl Acad. Sci. USA*, **95**, 11476–11481.
34. Johns, G. C. & Somero, G. N. (2004). Evolutionary convergence in adaptation of proteins to temperature: A4-lactate dehydrogenases of Pacific damselfishes. *Mol. Biol. Evol.* **2**, 314–320.
35. Girard, E., Marchal, S., Perez, J., Finet, S., Kahn, R., Fourme, R. *et al.* (2010). Structure–function perturbation and dissociation of tetrameric urate oxidase by high hydrostatic pressure. *Biophys. J.* **98**, 2365–2373.
36. Thomas, A. S. & Elcock, A. H. (2006). Direct observation of salt effects on molecular interactions through explicit-solvent molecular dynamics simulations: differential effects on electrostatic and hydrophobic interactions and comparisons to Poisson–Boltzmann theory. *J. Am. Chem. Soc.* **128**, 7796–7806.
37. Collins, T., Meuwis, M. A., Gerday, C. & Feller, G. (2003). Activity, stability and flexibility in glycosidases adapted to extreme thermal environments. *J. Mol. Biol.* **328**, 419–428.
38. Georlette, D., Damien, B., Blaise, V., Depiereux, E., Uversky, V. N., Gerday, C. & Feller, G. (2003). Structural and functional adaptations to extreme temperatures in psychrophilic, mesophilic, and thermophilic DNA ligases. *J. Biol. Chem.* **278**, 37015–37023.
39. Bae, E. & Phillips, G. N., Jr. (2004). Structures and analysis of highly homologous psychrophilic, mesophilic, and thermophilic adenylate kinases. *J. Biol. Chem.* **279**, 28202–28208.
40. Madern, D., Ebel, C., Dale, H. A., Lien, T., Steen, I. H., Birkeland, N. K. & Zaccai, G. (2001). Differences in the oligomeric states of the LDH-like L-MalDH from the hyperthermophilic archaea *Methanococcus jannaschii* and *Archaeoglobus fulgidus*. *Biochemistry*, **40**, 10310–10316.
41. Okemefuna, A. I., Nan, R., Gor, J. & Perkins, S. J. (2009). Electrostatic interactions contribute to the folded-back conformation of wild type human factor H. *J. Mol. Biol.* **391**, 98–118.
42. Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26**, 795–800.
43. Matthews, B. W. (1968). Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491–497.
44. McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. (2005). Likelihood-enhanced fast translation functions. *Acta Crystallogr., Sect. D.: Biol. Crystallogr.* **61**, 458–464.
45. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echoo, N. *et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr., Sect. D.: Biol. Crystallogr.* **66**, 213–221.
46. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, W. (2010). Features and development of Coot. *Acta Crystallogr., Sect. D.: Biol. Crystallogr.* **66**, 486–501.
47. Chen, V. B. W., Arendall, B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J. *et al.* (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr., Sect. D.: Biol. Crystallogr.* **66**, 12–21.
48. Eventoff, W., Rossmann, M. G., Taylor, S. S., Torff, H. J., Meyer, H., Keil, W. & Kiltz, H. H. (1977). Structural

Chapitre 8

AcrB, la protéine modèle membranaire putative

Dans le cadre du "développement de nouveaux outils" pour le phasage de macromolécules biologiques, la recherche de protéines éclectiques en guise de modèle d'étude nous a conduit vers "l'Acriflavine resistance protein B" ou "AcrB" et ce, pour deux raisons. La première est que cette protéine a toutes les caractéristiques d'un modèle pouvant démontrer le pouvoir phasant des complexes de lanthanide sur les grands assemblages protéiques. Il s'agit d'une protéine non seulement homotrimèrique de 341 kDa mais aussi membranaire, elle apporte ainsi de nouvelles propriétés de surface pour l'étude du mode de fixation des complexes. Elle est facile à produire, à purifier, à cristalliser et sa structure est connue. La seconde raison est, qu'appartenant à la classe des protéines membranaires, elle constitue un modèle pour l'étude d'un nouveau complexe de lanthanide synthétisé dans le but de cibler ce type de protéine : l'acide 1,3-bis-phytanylamide éthylène-tri-amine triacétique ou "DTTA-BPA". Nous verrons dans ce court chapitre pourquoi AcrB s'est avéré un mauvais modèle d'étude en ne dépassant pas la première étape requise pour l'étude du mode de fixation des complexes de lanthanide qu'est une diffraction à une résolution raisonnable.

8.1 Les deux protagonistes de la courte histoire membranaire

8.1.1 La protéine membranaire "EcAcrB"

L'Acriflavine resistance protein B est une protéine homotrimérique ancrée dans la membrane interne (cytoplasmique) de la bactérie à Gram négatif *Escherichia Coli* (Okusu and Ma, 1996). Associée aux protéines AcrA et TolC, l'assemblage AcrAB-TolC constitue une pompe à efflux dont le rôle est le transfert direct (en évitant le périplasme et la membrane externe) de molécules hydrophobes, hydrophiles ou amphiphiles entre le cytoplasme et l'extérieur de la cellule (Fralick, 1996) et dont la force motrice est proton-dépendante (Zgurskaya and Nikaido, 2000) Ainsi, en étant à même de rejeter ce type de molécules, EcAcrB confère à la bactérie une résistance accrue à un grand nombre d'antibiotiques, d'où son nom (Nikaido and Zgurskaya, 2001). Mais cette protéine membranaire est aussi un contaminant bien connu des biochimistes des protéines membranaire recombinantes. D'une part, elle est exprimée par *E. coli*, bactérie très largement utilisée comme support d'expression pour protéines recombinantes, membranaires ou non. Et d'autre part, la protéine naturelle possède deux histidines très accessibles en

son extrémité C-terminale. Elles constituent une "étiquette histidine" suffisante pour accrocher la protéine aux colonnes d'affinité en résine de nickel dont l'utilisation est, elle aussi, très largement répandue. Cette protéine peut donc se retrouver facilement dans les extraits finaux purifiés et, cristallisant aussi très bien, peut conduire à la formation de cristaux qui ne sont pas ceux espérés (Veesler *et al.*, 2008; Glover *et al.*, 2011).

La première structure d'EcAcrB a été résolue par diffraction de rayons X à 3,0 Å de résolution par Klaas M. Pos et Kay Diederichs (Pos and Diederichs, 2002), à qui l'on doit aussi toute la biochimie et les premières descriptions structurales. La meilleure résolution ayant pu être obtenue à ce jour pour une structure d'EcAcrB est de 2,80 Å (Murakami *et al.*, 2006) **[Figure 8.1]**.

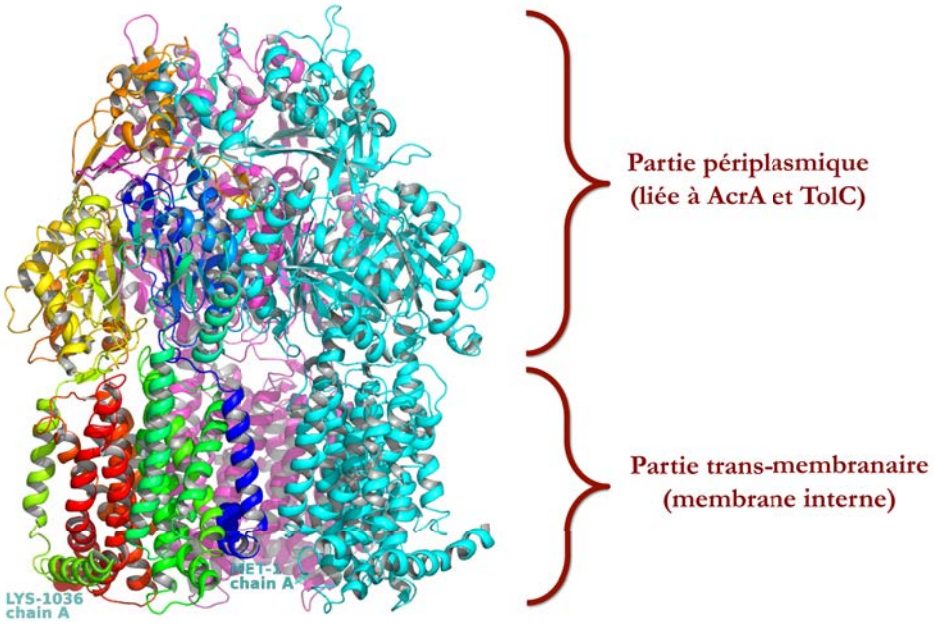


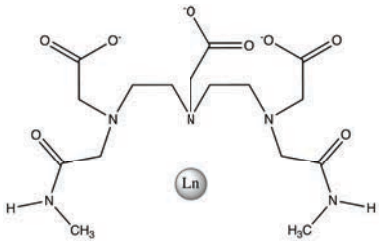
Figure 8.1 – EcAcrB, structure résolue à 2,80 Å. Le monomère A est en couleurs arc-en-ciel, de l'extrémité N-terminale (en bleu) à l'extrémité C-terminale (en rouge). Les monomères B et C sont colorés en cyan et magenta, respectivement.

L'analyse de la séquence primaire de la protéine EcAcrB révèle un poids moléculaire de 341 kDa (3147 acides aminés, 1049 par monomère), un point isoélectrique théorique de 5,39 et un coefficient d'extinction molaire mesuré est de 93000 $M^{-1}.cm^{-1}$.

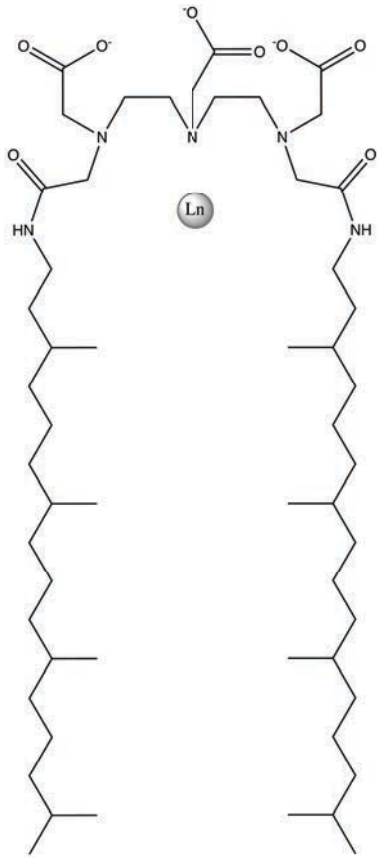
8.1.2 Le complexe de lanthanide DTTA-BPA

C'est dans le but de créer un complexe de lanthanide dont la structure cible spécifiquement les protéines membranaires que le complexe LnDTTPA-BPA a été synthétisé. Cette synthèse revient au Dr. Olivier Maury, du laboratoire de Chimie de l'Ecole Normale Supérieure de Lyon. Le complexe 1,3-bis-phytanylamide éthylène-tri-amine triacétate de lanthanide ou "LnDTTA-BPA" est dérivé du complexe LnDTPA-BMA. Deux chaines à seize carbones ramifiés ("2,6,10,14-tetraméthylhexadécane" ou "phytane") ont été greffées aux deux fonctions amides du ligand DTPA-BMA via leur groupement méthyle **[Figure 8.2]**. La structure finale du complexe présente ainsi une tête hydrophile quasi-identique au ligand DTPA-BMA qui lie l'atome de lanthanide et une queue très hydrophobe censée se lier à la partie

trans-membranaire des protéines membranaires. La ressemblance du complexe avec un phospholipide n'est pas un hasard : il a été synthétisé pour ressembler au "phytanyl", phospholipide qui compose la membrane des archées (Cavicchioli, 2007). Dérivé du complexe LnDTPA-BMA, le complexe LnDTTA-BPA ne peut émettre de fluorescence dans les longueurs d'onde du visible, la présence d'une molécule d'eau dans la sphère de coordination de l'atome de lanthanide venant la masquer. La charge globale du complexe de lanthanide est neutre. De formule chimique $C_{54}H_{102}N_5O_8{}^{3-}$, le ligand DTTA-BPA a une masse moléculaire de 948,42 g.mol^{-1}.



(a) Le complexe de lanthanide LnDTPA-BMA



(b) Le complexe de lanthanide LnDTTA-BPA

Figure 8.2 – Le complexe de lanthanide LnDTTA-BPA pour le phasage des protéines membranaires

Malheureusement, c'est tout ce que je puis dire à propos de ce complexe. Mes toutes premières tentatives pour solubiliser le complexe se sont révélées infructueuses et les premiers clichés de diffraction issus des cristaux natifs d'EcAcrB, enregistrés au même moments, n'ont guère été encourageants. Lors de premiers essais de solubilisation dans l'eau MilliQ, je n'ai pu évaluer qu'une concentration de seulement 3 µM environ avec le complexe EuDTTA-BPA (100 µg de complexe dans 30 mL d'eau MilliQ), et n'ai pas pu mesurer de concentration suite à une tentative de dilution dans une solution de détergent ß-DDM tant la solution était trouble. Je n'ai cependant pas essayé de diluer le complexe dans le DMSO par exemple, solvant organique polaire couramment utilisé en biochimie des protéines.

8.2 Production, purification et cristallisation de EcAcrB

8.2.1 Production & purification de EcAcrB

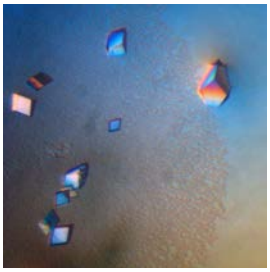
En annexe sont consignés les protocoles de production et de purification permettant d'obtenir de multiples cristaux de protéine EcAcrB parfaitement réguliers, pouvant avoisiner les 500 µm de longueur d'arête. Ils n'ont malheureusement jamais diffracté à une résolution supérieure à 6,5 Å. Mettre cependant par écrit les résultats obtenus de ces expériences de biochimie me permet de rendre compte de la qualité respective des contrôles de production, de cristallisation et de cryo-protection. Car cette qualité générale ne nous a jamais permis d'expliquer pourquoi les cristaux obtenus n'ont jamais diffracté à plus haute résolution, dénouement rendant le modèle EcAcrB inexploitable pour l'étude du pouvoir phasant des complexes de lanthanide.

Je tiens a remercier ici chaleureusement le Dr. Aurélien Deniaud, ami et actuellement post-doctorant dans la "team Schaffitzel", à l'E.M.B.L de Grenoble. En me formant à la production, la purification et à la cristallisation de la protéine EcAcrB avec didactisme, patience et exhaustivité, il m'a permis non-seulement d'utiliser la construction plasmidique de Klaas M. Pos mais aussi de m'introduire et de me former aux us et coutumes du laboratoire.

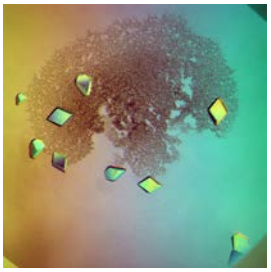
8.2.2 Cristallisation de la protéine EcAcrB et cryo-protection des cristaux

8.2.2.1 Cristallisation de la protéine EcAcrB

Tout comme la purification, la cristallisation de la protéine EcAcrB est aisée car elle cristallise dans de nombreuses conditions de cristallisation en goutte suspendue, conditions la plupart du temps à base de PEG. S'appuyant sur la littérature, nous avons retenu et amélioré les conditions décrites par Satoshi Murakami (Murakami *et al.*, 2002) qui conduisent à de très beaux cristaux d'EcAcrB **[Figure 8.3]**.



(a) Cristaux natifs de 6 mg.mL$^{-1}{}_{t_0}$ EcAcrB, 120 x 100 µm pour le plus gros



(b) Cristaux natifs de 6 mg.mL$^{-1}{}_{t_0}$ EcAcrB, 130 x 110 µm

5 - 20 % PEG 2000
40 mM Citrate de sodium pH 5,6
10 mM Phosphate de sodium pH 6,2
[P] = 12 mg.mL^{-1}
Goutte : 1,0 µL (P) + 1,0 µL (R)
T = 293 K (20°C)
48 heures

(c) Conditions de cristallisation

Figure 8.3 – Premières conditions de cristallisation d'EcAcrB utilisées. Le volume de solution dans les puits est ici de 500 µL

8.2.2.2 Cryo-protection des cristaux avant l'étape de diffraction des rayons X

Dans le processus qui conduit de la production de la protéine à la diffraction, la cryo-protection des cristaux natifs d'EcAcrB est l'étape la plus délicate. Bien que de tailles conséquentes, les cristaux sont très sensibles aux variations de concentrations lors des trempages et ils sont souvent piégés dans

un “filet élastique très résistant”, constitué de détergent et de protéine non-cristallisée rendant ainsi la pêche difficile **[Figure 8.3b]**.

La méthode de Klaas M. Pos s'est avérée la meilleure. Avant la vitrification en azote liquide, un cristal est trempé une première fois dans une solution identique à celle du puits avec 5 % de glycérol. Ce même cristal est ensuite trempé dans la solution cryo-protectrice finale de même composition que la précédente mais avec 30 % de glycérol. Dans le cas où les cristaux ne supportaient pas cet écart de concentration en glycérol, j'ai procédé de la même façon mais en multipliant les trempages intermédiaires par pas de 5 % de glycérol. Les temps de trempage varient de 2 à 5 minutes. Observés sous binoculaire, les cristaux d'EcAcrB semblent bien supporter l'une ou l'autre méthode.

8.3 Bref résumé des diverses tentatives d'amélioration

Malheureusement, aucun de ces cristaux au “faciès engageant” n'a jamais diffracté à une résolution supérieure à 6,5 Å.

Après un certain nombre d'essais, je me suis d'abord tourné vers la modification de l'étape de cryo-protéction, étape la plus délicate pouvant engendrer ces profils de diffraction inutiles à notre étude. Nous disposions à l'époque du générateur de rayons X de laboratoire pour tester rapidement l'effet des différentes modifications apportées au protocole de cryo-protection. Devant l'absence d'amélioration, je me suis alors concentré sur la cristallisation de la protéine pour remonter ensuite à l'étape de purification. Sans apporter de modification au protocole de production de la protéine EcAcrB. L'issue de cette étude n'ayant pas été un succès, il est inutile de détailler les différentes modifications apportées à chacune de ces étapes.

8.3.1 Résumé des modifications apportées au protocole de purification d'EcAcrB

Dans l'optique d'obtenir des cristaux d'EcAcrB diffractants à une résolution suffisante, je me suis permis de modifier le protocole de purification selon deux grands axes : le changement de détergent utilisé et l'adjonction de sel lors de la purification dans la solution de stockage de la protéine. L'autre détergent choisi est celui préconisé dans la littérature, le 6-Cyclohexyl beta-D-maltoside ou “CHM” (aussi connu sous le nom commercial de “Cymal-6”. Ainsi, quatre protocoles de purification, s'appuyant sur le protocole initial ont été menés de conser*ves* :

- ▷ Solubilisation et purification en présence de détergent ß-DDM, sans NaCl (identiques à celles décrites précédemment)
- ▷ Solubilisation et purification en détergent ß-DDM avec 200 mM NaCl
- ▷ Solubilisation et purification en détergent CHM sans NaCl
- ▷ Solubilisation et purification en détergent CHM avec 200 mM NaCl

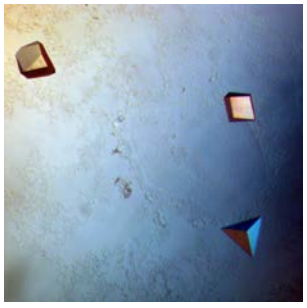
Comme nous le verrons dans la sous-section suivante, la protéine purifiée et stockée en présence de CHM a conduit à de plus beaux cristaux mais la solution de protéine pure semble moins stable à 4°C qu'avec le detergent ß-DDM. En effet, après trois semaines de conservation, un précipité blanchâtre sédimente au fond du tube de stockage. Bien que le chlorure de sodium ait été ajouté dans le but de stabiliser la protéine en solution et pour éventuellement modifier la croissance des futurs cristaux, aucun effet particulier n'a pu être constaté suite à cet ajout.

8.3.2 Résumé des tentatives d'amélioration des conditions de cristallisation et délit de faciès cristallin

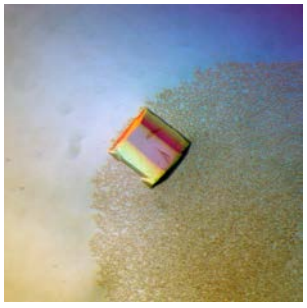
En plus de la condition de cristallisation transmise par le Dr. Aurélien Deniaud et Céline Julian-Binard, décrite ci-dessus, j'ai pu disposer de six conditions supplémentaires publiées à partir desquelles de nouvelles conditions ont pu être élaborées (Pos and Diederichs, 2002; Purdy *et al.*, 2002; Yu *et al.*, 2003; Pos *et al.*, 2004; Seeger *et al.*, 2006; Törnroth-Horsefield *et al.*, 2007). De plus, dans son second papier relatif à EcAcrB, Klaas M. Pos précise que ce ne sont pas forcément les cristaux de plus grandes tailles qui présentent le meilleur profil de diffraction mais les petits (Pos *et al.*, 2004).

J'ai pu obtenir des cristaux natifs d'EcAcrB en croisant des conditions de cristallisation pouvant contenir sept gammes de précipitant différentes (PEG 400, PEG 1500, PEG 2000, PEG 2000 MME, PEG 3350, PEG 4000, PEG 6000 et glycérol), quatre tampons différents (citrate de sodium pH 5,6, phospate de sodium pH 6,2, HEPES pH 7,0 et Tris HCl pH 8,0), trois gammes de sel différentes (chlorure de sodium, sulfate d'ammonium et sulfate de lithium) et avec ou sans 0,02 % de détergent. J'ai aussi fait varier le volume du puits (250, 500 et 1000 µL), le volume de la goutte (2, 3 ou 4,5 µl avec adjonction ou non d'un volume d'eau) et la solution stock concentrée en protéine (en présence de détergent ß-DDM ou CHM, avec ou sans NaCl).

En diminuant le nombre de points de nucléation dans des conditions où la diffusion de vapeur fut ralentie, des cristaux aux faces régulières dont les arêtes avoisinaient 400 µm de longueur ont pu être obtenus. Pour cela, la solution était concentrée en EcAcrB à 10 mg.mL^{-1}, purifiée avec le détergent CHM, dans des gouttes où 1 µL d'eau MilliQ a été ajouté, face à un volume de puits de 500 µL et dans les condition décrites dans la **figure 8.4**. La présence de glycérol permet au cristal de mieux résister au protocole de cryo-protection retenu.



(a) Cristaux natifs de d'EcAcrB, 230 x 300 µm



(b) Cristal natif d'EcAcrB, 410 x 380 µm

4 - 10 % PEG 4000
5 % glycérol
100 mM HEPES pH 7,0
100 mM NaCl
[P] = 10 mg.mL^{-1}
Goutte : 1,0 µL (P) + 1,0 µL (R) + 1 µL eau
T = 293 K (20°C)
3 - 5 jours

(c) Conditions de cristallisation

Figure 8.4 – Conditions de cristallisation conduisant aux cristaux d'EcAcrB offrant la plus grande taille et la plus grande régularité. Puits de 500 µL. Condition de purification et de stockage avec détergent CHM

A nouveau et malgré les différentes tentatives, aucun cristal, même de petite taille, n'aura diffracté à une résolution plus élevée que 6,5 Å. De plus, la solution de stockage en détergent CHM semblait mal vieillir à 4°C. La **figure 8.4** et l'information quant à la possibilité d'obtenir de meilleurs profils de diffraction avec les petits cristaux illustrent bien le dilemme auquel je me suis confronté (ou le piège dans lequel je suis tombé) lors de cette étude. Si "améliorer des conditions de cristallisation" revient à "rechercher les conditions de cristallisation conduisant aux plus beaux cristaux" alors il faut, dans le cas

d'EcAcrB, redéfinir la notion de "beau cristal". Est alors un "beau cristal" celui qui diffracte à haute résolution et non-pas celui qui présente le plus joli faciès cristallin. Dans cette mesure, même avec un accès privilégié à un générateur de rayons X de laboratoire, il devient difficile "d'améliorer", *de visu*, des conditions de cristallisation. Fus-je coupable de délit de faciès cristallin ?

Une fois seulement, un cristal natif d'EcAcrB a diffracté à 3,6 Å de résolution. Il s'agit du cristal présenté dans la **figure 8.4b**. Mais son profil de diffraction ne permit pas d'enregistrer la moindre donnée, tant le cristal avait souffert lors de l'étape de cryo-protection.

8.3.3 Résumé des tentatives d'amélioration du protocole de cryo-protection des cristaux d'EcAcrB

Toujours en s'appuyant sur la littérature mentionnée précédemment et sur les cryo-protectants utilisés au laboratoire, différents tests de diffraction ont été menés avec douze conditions de cryo-protection différentes. Les cryo-protectants candidats ont été le PEG 400, le PEG 2000, l'éthylène-glycol, le glycérol, la paraffine et la paratone. Ce qui semblait la meilleure combinaison obtenue était la cryo-protection des cristaux obtenus en présence de 5 % de glycérol (voir les conditions ci-dessus) réalisée par trempage dans des solutions formant une gamme de 5 à 30 % de glycérol, par pas de 5 %. Mais encore et toujours, bien que les cristaux semblaient parfaitement résister, aucun profil de diffraction obtenu ne fut digne d'enregistrement.

8.4 Conclusions quant à l'utilisation de la protéine modèle EcAcrB

Sur une quarantaine de cristaux testés sur la ligne de lumière Proxima1 du centre de rayonnement synchrotron SOLEIL et une quarantaine sur la ligne de lumière BM30A de l'E.S.R.F, une seule et unique fois, un cristal natif de protéine EcAcrB a diffracté à 3,6 Å de résolution. Résolution qui aurait pu être suffisante pour tester le pouvoir phasant des complexes sur un grand assemblage membranaire mais insuffisante pour étudier le mode de fixation de ces complexes sur une telle protéine. Le profil de diffraction fut de toute façon non-enregistrable tant ce cristal avait souffert. Devant cette maigre preuve du possible, nous nous sommes alors demandés si la solution n'était pas finalement d'amener une centaine de cristaux d'EcAcrB sur une ligne de lumière puis de tester leur diffraction à l'aide du chargeur automatique d'échantillons jusqu'à découvrir le cristal élu : le cristal d'EcAcrB diffractant à haute résolution.

Alors que d'autres projets présentaient déjà des résultats prometteurs (les projets "MalDH", "dipicolinate" ou encore le projet "PhTET1"), les difficultés rencontrées pour solubiliser le complexe de lanthanide DTTA-BPA et pour obtenir des cristaux de protéine modèle diffractant à haute résolution ont fait que cette étude a été mise en suspend. Cependant, il n'est pas exclu de trouver une autre protéine membranaire, modèle pour l'étude du pouvoir phasant et du mode de fixation des complexes de lanthanide en notre possession. Mais celle-ci devra regrouper les caractéristiques habituelles des protéines modèles : facile à produire, facile à purifier, d'une cristallisation rapide conduisant à des cristaux diffractant à haute résolution. Pour une protéine soluble, le choix est vaste mais il se rréduit pour les protéines membranaires. Après quelques recherches au sein de l'institut et après avoir mené une seule fois sa production et sa purification, il est possible que la protéine membranaire FhuA de *E.coli* constitue une bonne candidate (Ferguson *et al.*, 1998).

Conclusion quant à l'utilisation des complexes de lanthanide pour le phasage de macromolécules biologiques de grande taille

Les études préliminaires sur le complexe DTTA-BPA, nouveau complexe dont le mode de fixation putatif est de cibler les parties transmembranaires des protéines membranaires se sont avérées plus difficiles qu'escomptées. D'une part, les essais de solubilisation du complexe n'ont pas abouti à une solution suffisamment concentrée en complexe et d'autre part la protéine choisie, EcAcrB, ne s'est pas pas révélée un bon modèle. En effet, bien que facile à produire, à purifier et à cristalliser, aucun cristal natif d'EcAcrB n'a diffracté à une résolution supérieure à 6,5 Å. Or la haute résolution nous est indispensable pour déterminer le mode de fixation subséquent des complexes. Cela est d'autant plus dommageable que la protéine entrait aussi dans la catégorie des grandes protéines multimériques (homotrimère de 341 kDa). Elle aurait ainsi pu constituer un bon représentant pour illustrer le pouvoir phasant des complexes de lanthanide. Si dans le cas d'EcAcrB la solution était d'amener une centaine de cristaux sur une ligne de lumière afin d'isoler celui qui diffracte à haute résolution, approche incompatible avec une étude systématique des complexes de lanthanides, nous avons préféré nous tourner vers une autre protéine modèle. C'est pourquoi nous avons envisagé de poursuivre à l'avenir cette étude avec la protéine membranaire EcFhua dont les protocoles de production, purification et de cristallisation nous ont été généreusement transmis par le Dr. Cécile Breyton, de l'équipe SSIMPA du groupe "Membrane et pathogène" de l'institut.

A l'inverse, l'étude sur l'utilisation des complexes de lanthanide pour le phasage de grands assemblages protéiques s'est avérée plus riche en résultats.

Nous avons montré que le complexe tris-dipicolinate d'europium, $[Eu(DPA)_3]^{3-}$ a permis de déterminer la structure de l'amino-peptidase homododécamérique tétraédrique PhTET1-12s, malgré une résolution minimale des données de 4,0 Å, Le phasage par la méthode SAD a été réalisé avec le jeu de données anomales issu du cristal dérivé $[Eu(DPA)_3]^{3-}$ à 4,0 Å de résolution. La carte de densité électronique expérimentale obtenue s'est avérée parfaitement interprétable pour cette tranche de résolution, prouvant que les complexes de lanthanide peuvent conduire à un phasage précis même à basse résolution.

Les complexes de lanthanides issus de l'imagerie médicale ont aussi permis déterminer la structure de trois nouvelles enzymes homotétramériques : la malate déshydrogénase de la bactérie thermophile

Chloroflexus aurantiacus "CaMalDH", la malate déshydrogénase de l'archée hyperthermoacidophile *Picrophilus torridus* "PtMalDH" et la malate déshydrogénase de l'archée halophile *Haloferax volcanii* "HvMalDH", à respectivement 1,7 Å, 2,0 Å et 1,5 Å de résolution. La détermination de ces structures à haute résolution par les méthodes de phasage *de novo* a permis la modélisation de trois éléments structuraux particuliers qui apportent de nouveaux éclairages sur l'adaptation halophile.

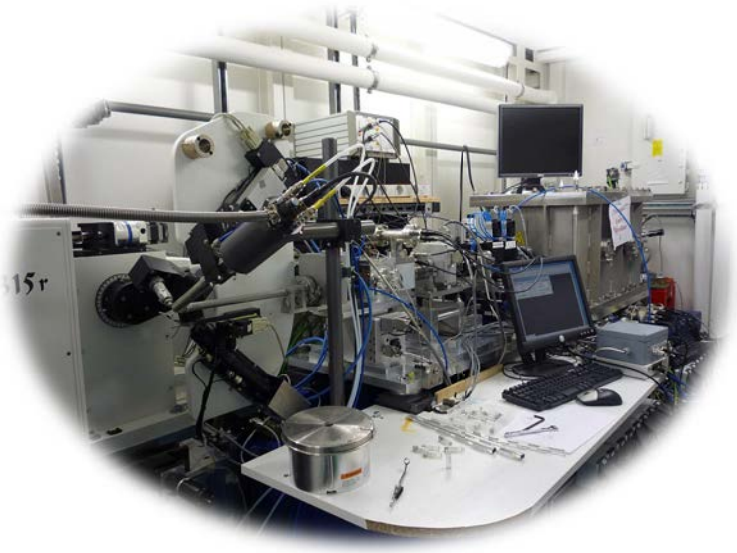
En premier lieu, nous avons pu identifier et positionner les ions Cl^- et K^+ au sein de la structure de HvMalDH. Le positionnement précis de ces ions permet d'étudier les caractéristiques structurales de ces sites ioniques potentiels et de regarder leur conservation phylostructurale par études comparatives avec d'autres structures de MalDH et LacDH halophiles. Ce qui permettra d'analyser l'influence des sites ioniques clés sur la structure de ces enzymes halophiles.

Deuxièmement, en lien avec l'activité enzymatique, plusieurs conformations de la boucle catalytique du site actif de PtMalDH ont pu être modélisées. Il convient de noter que c'est la première fois qu'il est possible d'observer par cristallographie les changements conformationnels associés à la catalyse d'une malate déshydrogénase tétramérique. Les réorganisations conformationnelles n'avaient été décrites jusqu'alors que pour les LacDH. Bien que ce ne soit pas illustré de manière détaillée dans ce manuscrit, la comparaison des formes ouvertes et fermées entre malate et lactate déshydrogénases montrent des différences notables qui seront exploitées ultérieurement dans le cadre d'analyses phylostructurales.

Troisièmement, nous avons pu mettre en évidence une organisation singulière des molécules d'eau à la surface de CaMalDH. Les protéines halophiles ont principalement évolué en modifiant leurs surfaces d'interaction avec le solvant, ce qui engendre une modification de leur enveloppe de solvatation, en solution, par rapport à leurs homologues non halophiles. Grâce à la détermination de la structure de l'enzyme CaMalDH à haute résolution, il a été possible de démontrer pour la première fois que la minimisation de l'exposition au solvant de groupements hydrophobes à haut sel, pourtant défavorable d'un point de vue énergétique, est la principale force évolutive qui dicte l'adaptation des protéines halophiles.

Ces enzymes MalDH partagent plus de 72 % de similarité de séquence et présentent pourtant des propriétés de surface très différentes en fonction de leur caractère adaptatif. Ainsi, en utilisant ces enzymes comme bibliothèque de fonctions chimiques, nous avons pu mettre en place une nouvelle approche méthodologique pour comprendre finement les modes d'interaction des complexes de lanthanides. Cette approche a été mise à l'épreuve avec les trois malate déshydrogénases différentes précitées et trois complexes IRM : GdHPDO3A, GdDO3A et GdDTPA-BMA. Les résultats préliminaires obtenus par cette étude indiquent d'ores et déjà des conditions et des modes de fixation qui diffèrent pour les trois complexes IRM testés selon la composition des surfaces des protéines. En comparant les propriétés électrostatiques de la région du premier site de fixation du complexe GdHPDO3A sur CaMaDH avec celles de la région correspondante chez PtMalDH et HvMalDH, nous avons confirmé que la fixation du complexe GdHPDO3A requiert une zone hydrophobe mais pas nécessairement très marquée. Cependant, ces observations ont aussi mis en évidence un facteur de forme de surface qui semble avoir une importance dans la fixation de ce complexe. Dans la région du premier site de CaMaDH où le complexe se fixe, la topologie de la surface semble faire apparaitre une proéminence hydrophobe sur laquelle le complexe vient se fixer. Dans la région correspondante chez PtMalDH, où le complexe ne vient pas se fixer alors qu'il vient se fixer à d'autres sites sur cette protéine, on ne retrouve pas cette aspérité au même endroit que la zone hydrophobe. D'autre part, cette approche a aussi permis de confirmer le mode de fixation

du complexe GdDO3A. A la surface de HvMalDH, riche en acides aminés aspartate et glutamate, il se fixe non pas par l'intermédiaire du ligand mais en créant une liaison ionique entre l'ion lanthanide et la fonction acide de ces acides aminés. En outre dans le cadre de cette approche, nous avons aussi comparé une région de la structure de HvMalDH où vient se fixer un complexe GdDO3A avec la région équivalente sur la structure de HvMalDH issue d'un cristal natif. Cette comparaison a révélé une probable compétition entre une molécule de PEG et le complexe qui peut se produire lors d'une expérience de trempage. Au regard des résultats préliminaires obtenus, nous pouvons espérer qu'à l'issue de ce travail, nous pourrons déterminer quel sera le meilleur complexe en fonction de la séquence en acides aminés de la protéine-cible et des conditions de cristallisation.



Discussion et perspectives

Tout d'abord, au cours de cette discussion, je ne reviendrai ni sur la détermination des structures de malate déshydrogénases, ni sur les hypothèses qu'elles ont permis d'émettre quant à l'adaptation halophile (l'importance des sites ioniques et de l'organisation du solvant). Je vais donc me focaliser sur les complexes de lanthanide et leur utilisation.

Au cours de ce travail de recherche, nous avons mené une étude sur un nouveau complexe luminescent, le triazonanetripicolinate de lanthanide, et nous avons poursuivi celles entreprises précédemment au sein de l'équipe. Le développement de ce nouveau complexe a permis d'enrichir la collaboration avec le Dr. Olivier Maury du Laboratoire de Chimie de l'Ecole Normale Supérieure de Lyon, collaboration déjà établie avant mon arrivée. Nous avons aussi participé au développement et à l'étude d'une nouvelle famille de complexes luminescents, les tris-triazoledipicolinates de lanthanide. Ce développement m'a conduit à établir une nouvelle collaboration avec Lionel Nauton et Arnaud Gautier du laboratoire SEESIB de l'Université Blaise-Pascal de Clermont-Ferrand. Ces nouveaux complexes ont permis le phasage *de novo* des deux protéines couramment utilisées comme modèles dans le domaine de la cristallographie : le lysozyme de blanc d'œuf de poule et la thaumatine de *Thaumatococcus daniellii*.

Depuis quatorze années, un nombre important de structures de protéines, de tailles et de propriétés de surface variées, a été résolu par phasage expérimental grâce aux complexes de lanthanide. Rien qu'au cours des travaux réalisés pour cette thèse, ont ainsi été résolues par phasage *de novo* : la structure du lysozyme de blanc d'œuf de poule (17 modèles), la structure de la thaumatine de *Thaumatococcus daniellii* (4 modèles), la structure de CaMalDH (3 modèles), la structure de PtMalDH (2 modèles), la structure de HvMalDH (3 modèles) et la structure de PhTET1-12s (1 modèle).

En résolvant les structures de grandes protéines multimériques par les méthodes de phasage exploitant le signal anomal des lanthanides, nous avons confirmé leur haut pouvoir phasant et ce, même avec des données à basse résolution. De plus, les structures des trois malate déshydrogénases déterminées au cours de cette thèse montrent qu'au moins un complexe de lanthanide a pu se fixer. Ce constat confirme notre intuition sur le fait que le nombre de site de fixation adéquat à la surface des protéines est d'autant plus élevé que la taille de la protéine est importante.

Au début de mon travail de recherche, je m'étais promis qu'en fin de thèse, je pourrais répondre avec certitude à LA question : "Quel complexe dois-je utiliser pour résoudre la structure de ma protéine ?"

A l'heure actuelle, je ne peux répondre qu'avec circonspection à cette question, tant les paramètres dont il faut tenir compte sont nombreux. Cependant, on peut :

- ▷ préconiser l'utilisation du complexe **LnDO3A** si la surface de la protéine est très acide ou si la séquence primaire présente un nombre important d'acides aminés aspartate ou glutamate. Il est utilisable en trempage ou en co-cristallisation dans n'importe quelle condition de cristallisation. Aussi, faut-il l'utiliser à haute concentration (à plus de 80 mM).
- ▷ conseiller le complexe $\mathbf{[Ln(HMTDPA)_3]^{3-}}$ si la séquence primaire présente un nombre important d'acides aminés tryptophane et si les conditions de cristallisation n'impliquent pas l'utilisation de sels d'ions divalents et/ou de fortes concentrations en sels. De plus, ce complexe étant luminescent dans le visible, une lampe UV de paillasse pourra facilement révéler sa fixation dans les cristaux.

C'est le complexe qui présente la plus forte affinité pour les macromolécules biologiques, il pourra donc être utilisé à faible concentration (de 3 à 10 mM). Par contre, il vaut mieux utiliser le $[Ln(HMTDPA)_3]^{3-}$ en co-cristallisation car sa fixation semble moins évidente par trempage.

- ▷ recommander le **LnHPDO3A** si les conditions de cristallisation ne permettent pas l'utilisation du $[Ln(HMTDPA)_3]^{3-}$. Le LnHPDO3A est aussi connu pour établir des interactions hydrophobes avec les acides aminés tryptophane. Il n'est pas luminescent mais il est tout à fait utilisable en trempage. Il est stable quelle que soit la condition de cristallisation et, comme le LnDO3A, il vaut mieux l'utiliser en concentration saturante (plus de 80 mM).
- ▷ préconiser le complexe **$[Ln(DPA)_3]^{3-}$** si la solubilisation de la protéine d'intérêt semble sous-entendre une surface protéique très basique ou si la séquence primaire présente une importante quantité d'acides aminés arginine ou lysine. Tout comme le $[Ln(HMTDPA)_3]^{3-}$, il a l'avantage d'être luminescent, on peut donc détecter sa fixation dans les cristaux. Mais il reste inutilisable si les conditions de cristallisation de la protéine exigent l'utilisation de sels d'ions divalents et/ou de fortes concentrations en sels. Il faudra l'introduire par co-cristallisation en l'utilisant à plus de 50 mM car sa fixation ne semble pas suffisante si on l'utilise à une concentration plus basse.
- ▷ recommander le complexe **LnDTPA-BMA** en guise d'alternative au $[Ln(DPA)_3]^{3-}$. Son mode de fixation semble privilégier les liaisons hydrogènes avec les acides aminés lysine, arginine, glutamine et thréonine et, contrairement au $[Ln(DPA)_3]^{3-}$, reste stable en présence de sels d'ions divalents ou de fortes concentrations en sels. Le LnDTPA-BMA présente aussi l'avantage d'être utilisable en trempage. Mais quelle que soit la méthode d'obtention de cristaux dérivés, il faut l'utiliser à une concentration d'au moins 80 mM.

Enfin, si aucun choix préalable ne peut être fait, **il est toujours possible de les utiliser conjointement**. Au cours de mon travail de thèse, je n'ai pas constaté quoi que ce soit qui puisse contre-indiquer une telle méthode. Les complexes ne semblent pas interagir entre eux et le seul risque auquel on s'expose est d'avoir une compétition entre deux complexes si ceux-ci partagent le même site de fixation. Mais cela n'empêchera pas le phasage ultérieur si les complexes se fixent.

Les différentes études menées permettent de définir les axes d'analyse nécessaires à la compréhension fine des modes d'interaction. Ainsi la fixation des complexes de lanthanide dépendra de :

- ▷ La séquence en acides aminés de la protéine, en particulier la proportion de certains types de résidus
- ▷ La répartition des charges électrostatiques à la surface de la protéine
- ▷ La topologie de la surface de la protéine
- ▷ La composition chimique et du pH de la solution de cristallisation qui moduleront la fixation des complexes.

L'approche basée sur la famille des malate déshydrogénases devrait permettre de comprendre le poids et l'influence de chacun des points présentés ci-dessus. En effet, si le premier et le dernier points sont connus pour une protéine en cours d'étude structurale, il n'en est pas de même pour la répartition de la charge ou pour la topologie qui suppose que la structure est connue. Ainsi pour chaque structure déterminée, y compris les structures natives, il faudra regarder la topologie des surfaces, en particulier si la zone d'interaction fait partie d'une hélice, d'un brin bêta ou simplement d'une boucle (analyse qui finalement n'a jamais été entreprise jusqu'à présent) et cartographier la répartition des charges et ainsi déterminer l'importance pour l'interaction d'un certain type de résidus. L'étude fine du mode de fixation

des complexes dans leur site le plus peuplé pourra alors commencer afin de déterminer les résidus participant à l'interaction. Cela impliquera, pour chaque complexe, la comparaison site à site avec les régions équivalentes de toutes les MalDHs dont les structures seront disponibles. Il est à noter que l'équipe à déjà en main la production et la purification d'autres MalDHs aux propriétés de surface probablement différentes, comme la MalDH de l'archée alkalophile halophile *Natrialba magadii*, celle de la bacterie piézothermophile *Thermoaerobacter marianensis* ou encore celle de la bactérie hyperacidophile *Methylacidiphilum infernorum*. Cette analyse et celle des sites secondaires permettront ensuite de déterminer si les résidus en amont et en aval de l'acide aminé impliqué principalement dans l'interaction module la fixation et de quelle manière. Enfin l'influence de la composition de la solution de cristallisation pourra être corrélée au site de fixation. Par exemple, la présence de forte concentration en sels renforcera les interactions de type hydrophobe alors que ces mêmes interactions seront atténuées en présence de PEG. La possibilité de disposer de plusieurs conditions de cristallisation différentes, comme c'est souvent le cas pour les MalDHs, permettra de déterminer finement l'influence de la composition de la solution de cristallisation.

A titre d'exemple, nous avons vu l'importance des liaisons hydrogènes et des liaisons ioniques pour la fixation des complexes : liaisons hydrogènes entre la fonction acide des ligands dipicolinates et la fonction amine des acides aminés arginine et lysine, liaisons ioniques entre l'ion lanthanide et le groupement acide des acides aminés aspartate et glutamate. Mais tous ces groupements protéiques ionisables ont un pKa en solution La protonation de ces acides aminés est toujours dépendante du pH de la solution tampon utilisée dans les conditions de cristallisation. De fait, la fixation des complexes doit être dépendante de cette protonation. Que se passe-t-il si, à très bas pH, les acides aminés aspartate et glutamate sont protonés ? Pourront-t-il toujours fixer l'ion lanthanide du complexe ? Si les acides aminés arginine et lysine sont déprotonés à pH très basique, pourront-t-ils établir des liaisons hydrogènes avec les ligands des complexes ? L'influence du pH pourrait, par exemple, expliquer l'absence totale de fixation du complexe GdDO3A sur l'enzyme PtMalDH. En effet, les cristaux sont obtenus à pH 3,5. A ce pH, tous les groupements acides des acides aminés aspartate et des glutamate doivent être protonés. Ainsi, de part leur état carboxylique et non plus carboxylate, il est possible que ces fonctions ne puissent plus s'impliquer dans une liaison ionique avec l'ion lanthanide du complexe LnDO3A.

En fonction des résultats qui seront obtenus sur une série large de MalDHs, nous pourrons proposer une méthode d'analyse de la séquence d'une protéine de structure inconnue et ainsi répondre à la question "*Quel complexe dois-je utiliser pour résoudre la structure de ma protéine ?*". D'ores et déjà, je peux anticiper que cette méthode aura pour but de déterminer la présence, sur la séquence, de motifs adéquats à la fixation des complexes. Pour cela, sur la base d'une séquence protéique, nous pourrons, à l'aide de programmes de détermination de structures secondaires, déterminer les secteurs présentant le plus de chance d'accueillir tel ou tel complexe (boucle par exemple). Puis au sein de ces secteurs, vérifier la présence des résidus propices à l'interaction et enfin appliquer une pénalité liée aux conditions de cristallisation. Comme résumé sur la **figure 8.5**, nous pourrons ainsi estimer quel est le complexe le plus propice pour déterminer la structure de la protéine.

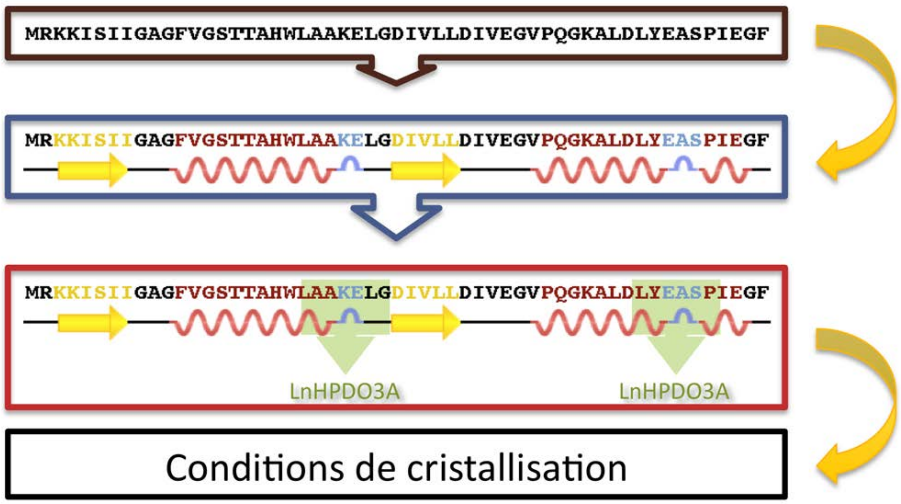


Figure 8.5 – Le but à atteindre : pouvoir choisir le meilleur complexe en fonction de la protéine à étudier

En ce qui concerne la compréhension de l'effet supramoléculaire observé dans le cas du complexe $[Ln(DPA)_3]^{3-}$, l'étude que nous venons d'entreprendre en comparant les structures des protéines HEWL et TEWL, auxquelles le complexe $[Ln(DPA)_3]^{3-}$ s'est fixé, pourra apporter un nouvel éclairage. Cette étude pourra être combinée aux résultats obtenus sur la combinatoire des ligands dans le cas des complexes $[Ln(TDPA)_3]^{x}$.

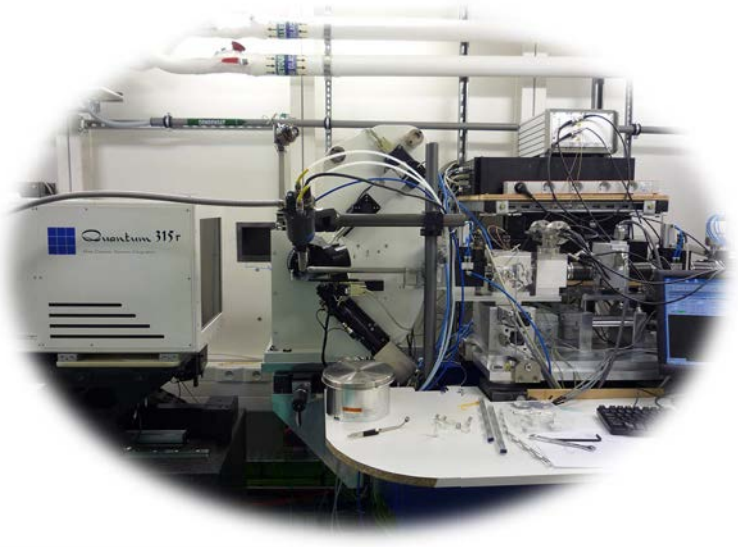
La quête du complexe parfait

Disposer d'un complexe de lanthanide unique très soluble, capable de se fixer au plus grand nombre de protéines, utilisable dans n'importe quelle condition de cristallisation et dont la fixation dans les cristaux peut être détectée par une méthode simple est notre but ultime. A l'issue de ce travail de thèse, je souhaite toutefois formuler une sorte de mise en garde. D'une part, nous avons pu constater qu'à chaque nouveau complexe synthétisé, pour un certain nombre de nouvelles propriétés gagnées, on perd une ou plusieurs propriétés désirables (ou on gagne des indésirables). D'autre part, nous avons encore en notre possession un certain nombre de complexes de lanthanide lyophilisés, produits par la société Bracco Imaging®, que nous n'avons pas encore pu étudier. Je pense donc qu'il ne faut pas se fourvoyer en s'aventurant trop tôt dans la synthèse et l'analyse d'un tout nouveau complexe de lanthanide. Il faut en premier lieu parfaire notre connaissance des modes de fixation des complexes utilisés jusqu'à présent, ce que les études MalDH et HEWL/TEWL devraient permettre.

En l'état actuel de nos connaissances, le complexe quasi idéal est certainement le LnDO3A. En effet, il présente un mode de fixation privilégiant les très ubiquistes acides aminés aspartate et glutamate et une souplesse d'utilisation (en co-cristallisation ou en trempage, quelles que soient les conditions de cristallisation). Toutefois il lui manque la propriété d'être facilement détectable dans les cristaux dérivés. Sur la base des différentes thèses effectuées sur cette thématique au sein de l'équipe, le LnDO3A est le complexe qui présente le taux de succès en phasage le plus élevé. Même si les taux d'occupation ne sont pas toujours les plus élevés, il conduit un nombre important de sites de fixation, en adéquation avec son mode d'interaction, qui assurent la réussite du phasage. Il est à noter que c'est le seul complexe

qui présente une coordination incomplète de l'ion lanthanide avec la présence de deux molécules d'eau dans la première sphère, ce qui explique son mode d'interaction. Si on utilise ce complexe comme point de départ pour une amélioration future, il faudra que la modification laisse la coordination incomplète, point qui risque d'être difficile si l'on souhaite, par exemple, obtenir une luminescence dans le visible.

Une autre voie pourrait être le développement d'une étiquette protéique pouvant être ajoutée à la protéine et ciblant tel ou tel complexe de lanthanide. Biochimiste de formation, je sais que cette méthode peut être longue à mettre en oeuvre et présente un caractère aléatoire. En effet il pourrait être difficile de préparer une telle étiquette par les méthodes de biologie moléculaire qui implique le choix d'un grand nombre de paramètres pour la mise en place du protocole (conception des amorces en fonction des températures d'hybridation et des codons rares, choix du vecteur d'expression, choix de l'organisme d'expression, choix de la méthode de transformation, choix de la polymérase, choix de la durée et de la température pour chacune des étapes PCR, choix des tampons, etc...). Cette investigation peut prendre un certain temps, elle est souvent tributaire du déroulement des premières expériences. Il se pose ensuite le problème de l'insertion même de cette étiquette dans la protéine-cible. En effet, le choix le plus évident est d'insérer la séquence d'ADN codant l'étiquette afin que celle-ci, une fois transcrite puis traduite, soit à l'extrémité N-terminale ou C-terminale de la protéine. Or, si on s'appuie sur les structures disponibles de la Protein Data Bank, on constate qu'il est très fréquent que les extrémités N-terminales et C-terminales se replient toutes deux vers l'intérieur de la protéine, parfois au point d'interagir ensemble. De plus, la lecture des articles du groupe de B. Imperiali (Silvaggi *et al.*, 2007) sur le lanthanide-binding-tag montre la difficulté de ce type d'approche. Par conséquent, développer une étiquette riche en acides aminés aspartate ou glutamate ciblant le complexe LnDO3A ne me paraît pas une bonne voie car trop couteuse en terme de développement comparé au gain que l'on peut escompter. En revanche, l'investissement dans le développement d'une étiquette riche en arginine ou en lysine serait acceptable dans le cadre de l'utilisation du $[Ln(DPA)_3]^{3-}$ en tant qu'agent de cristallisation en mettant à profit ses propriétés supramoléculaires.



Quatrième partie

Annexes

Chapitre 9

Séparation des protéines par électrophorèse sur gel de polyacrylamide en présence de dodécylsulfate de sodium

Je transcris ici le protocole que j'ai utilisé pour réaliser l'ensemble des colorations de protéines sur gel de polyacrylamide durant ce travail de thèse. Je n'ai utilisé que des gels dits dénaturant (en présence du détergent DodécylSulfate de Sodium "SDS").

9.1 Préparation des gels dénaturants de polyacrylamide

Dans le **tableau 9.1** qui suit, je donne la composition des gels dénaturants réalisés au cours de ces travaux de thèse. Le pourcentage d'acrylamide requis est donné en fonction de la résolution du gel requise (indiquée en gamme de poids moléculaires pouvant être séparés). Au laboratoire, nous utilisons le mélange Laemmli, préparé pour la communauté, qui contient déjà le tampon Tris HCl et le SDS nécessaire à la préparation des gels de polyacrylamide. Le mélange Laemmli est préparé avec 25 mM de Tris Base, 192 mM de Glycine et 0,1 % SDS.

Running	**6 %**	***8 %***	**10 %**	***12 %***	**15 %**	**20 %**	**Stacking**
	55-200 kDa	***30-90 kDa***	**16-68 kDa**	***14-56 kDa***	**12-43 kDa**	**? kDa**	
Eau MilliQ	5,3 mL	*4,6 mL*	4 mL	*3,3 mL*	2,3 mL	630 µL	2,7 mL
Acryl 30 %	2 mL	*2,66 mL*	3,33 mL	*4 mL*	5 mL	6,66 mL	670 µL
Tris / SDS	2,6 mL	*2,6 mL*	2,6 mL	*2,6 mL*	2,6 mL	2,6 mL	540 µL
APS 10 %	100 µL	*100 µL*	100 µL	*100 µL*	100 µL	100 µL	40 µL
Temed	8 µL	*6 µL*	4 µL	*4 µL*	4 µL	4 µL	4 µL

Tableau 9.1 – Composition pour préparation de 10 mL de gels dénaturants en fonction du pourcentage d'acrylamide souhaité. En italique, les deux concentrations en polyacrylamide utilisées au cours de cette thèse

9.2 Préparation du colorant au bleu de Coomassie

Pour colorer les protéines sur gel de polyacrylamide, un colorant au bleu de Coomassie a été systématiquement utilisé. Sa composition est dans le **tableau 9.2** qui suit.

Solutions et poudres stock	**[C]**$_{Finale}$	[C]$_{Stock}$	[Qté]$_{à\ prélever}$
Coomassie G-250	**0,2 mg.mL**$^{-1}$	-	10 mg
Tris HCl pH 6,8	**0,1 M**	1 M	5 mL
SDS	**8 %**	20 %	20 mL
Glycérol	**24 %**	100 %	12 mL
DTT	**0,2 M**	1 M	10 mL
Eau MilliQ	**q.s.p**	-	3 mL

Tableau 9.2 – Composition pour 50 mL de colorant "Sample Buffer 5X". **[C]**$_{Finale}$ est la concentration finale souhaitée, [C]$_{Stock}$ est la concentration des solutions ou des poudres stock et [Qté]$_{à\ prélever}$ correspond au volume à prélever ou à la masse à peser pour 50 mL de colorant

9.3 Marqueurs de poids moléculaires utilisés

Deux types de marqueurs de poids moléculaires ont été utilisés (à raison d'un dépot de 5 µL par gel). Ils proviennent tous deux des laboratoires Fermentas® et répondent aux noms de "*PageRuler™ Prestained Protein Ladder*" (référence SM0671) et *"PageRuler™ Unstained Protein Ladder*" (référence SM0661). Le premier n'a été utilisé que lors de la purification de la protéine membranaire AcrB, le second pour tous les autres gels, notamment ceux qui permirent de suivre la purification de la protéine AvEGFP.

Chapitre 10

Les protéines modèles

10.1 L'urate oxidase de *Aspergillus flavus* "AfUOx"

L'urate oxydase (ou encore "uricase") issue du champignon *Aspergillus flavus* est une enzyme appartenant à la famille des oxydoréductases (E.C. 1.7.3.3) qui catalyse l'oxydation de l'acide urique en 5-Hydroxyisourate. Chez l'homme, ce produit peut alors être à son tour oxydé en allantoïne, composé chimique utilisé aussi bien en médecine qu'en cosmétique, notamment en tant que produit cicatrisant et régénérant cellulaire. Cette enzyme est active sous la forme d'un homotétramère de 137 kDa (4x34,24 kDa) composé de 1204 acides aminés (4x301). Son point isoélectrique théorique est de 7.18 et son coefficient d'extinction molaire est de $\varepsilon_{280\,nm} = 53525\ M^{-1}.cm^{-1}$.

Un aliquot de 200 µL de solution comprenant la protéine concentrée à 16 $mg.mL^{-1}$ m'a été confié par le Dr. Nathalie Colloc'h de l'équipe CERVOxy du Laboratoire ISTCT de Caen et par le Pr. Thierry Prange de Laboratoire de Cristallographie et RMN Biologiques de l'Université Paris V René Descartes de Paris. Ils sont également auteurs des deux premières structures de l'AfUox (en présence ou en absence de l'inhibiteur 8-azaxanthin) obtenues en 2004 par diffraction des rayons X aux résolutions respectives de 1,75 Å et 2,30 Å. C'est initialement dans le but de trouver d'autres conditions de cristallisation pour la protéine, mais cette fois sans son inhibiteur (8-azaxanthine), qu'elle nous a été confiée. En effet, comme le but intrinsèque était d'obtenir des cristaux sans inhibiteur diffractant à une meilleure résolution et connaissant les propriétés supramoléculaire du $[Eu(DPA)_3]^{3-}$ pouvant conduire à de tels cristaux, nous nous sommes proposé de co-cristalliser l'AfUox avec ce complexe.

10.2 L'enhanced Green Fluorescent Protein de *Aequorea victoria* "AvEGFP"

10.2.1 Présentation de la protéine et données de biologie moléculaire

La Green Florescent Protein est une protéine monomerique soluble du cytoplasme de cellules de la méduse *Aequorea victoria*. La protéine se caractérise par sa structure en tonneau β qui abrite un chromophore qui la rend intrinsèquement fluorescente et photo-excitable. La protéine naturelle présente deux maxima d'excitation, aux longueurs d'ondes λ = 395 nm et λ = 475 nm et un maximum d'émission à la longueur d'onde λ = 504 nm, qui la rend aisément identifiabl e avec un spectrophotomètre.

Décrite pour la première fois en 1962, elle est actuellement souvent utilisée en tant que protéine de fusion dans le but de suivre l'extraction et la purification d'une protéine cible ou l'expression de cette

dernière dans son environnement naturel : la cellule vivante. Cette technique nécessite au préalable l'introduction du gène recombinant codant la protéine à étudier joint à celui de la GFP.

On doit la première structure de la GFP naturelle à M. Ormo et S. J. Remington en 1996, obtenue à une résolution de 1,90 Å par diffraction des rayons X.

Depuis, plusieurs variantes de la GFP ont pu être obtenue par mutagenèse dirigée pour modifier leur propriétés de fluorescence (fluorescence dans le cyan, le bleu ou le jaune) ou leur stabilité en solution. Lors de ce travail de thèse nous avons utilisé une de ces variantes : la Enhanced Green Fluorescent Protein ou "EGFP". Constituée de 238 acides aminés pour un poids moléculaire de 26,8 kDa et un point isoélectrique théorique de 5.58, cette protéine modifiée contient les mutations F64L et S65T qui, jouxtant le chromophore, provoquent le décalage du deuxième maximum d'absorption à $\lambda = 488$ nm et augmente l'émission de fluorescence de la protéine. Une troisième mutation H231L permet d'améliorer sa stabilité en solution tamponnée. Ses coefficients d'extinction molaire se retrouvent quelque peu modifiés par rapport à la protéine naturelle et prennent pour valeur : $\varepsilon_{280\,nm} = 22015\ \mathrm{M}^{-1}.50^{-1}$ et $\varepsilon_{488\,nm} = 63000\ \mathrm{M}^{-1}.\mathrm{cm}^{-1}$.

Un clone de la bactérie *Escherichia coli*, souche BL21(DE3), contenant un plasmide pRSET-B avec le gène recombinant codant la EGFP, une étiquette histidine et un site de clivage à l'α-chymotrypsine en N-terminal m'a été confié par notre collaborateur, le Dr. Philippe Oger du Laboratoire de Sciences de la Terre à l'Ecole Normale Supérieure de Lyon. A partir de ce clone, j'ai pu produire, purifier et cristalliser la protéine au laboratoire. La **figure 10.1** montre la séquence d'ADN insérée dans le plasmide pRSET-B. Aussi, ce plasmide confère à la bactérie *Escherichia coli* la résistance à l'antibiotique ampicilline. Des stocks en glycerol de EGFP ont été réalisés à partir d'aliquots d'une culture juste avant induction (DO ≈ 0,65) avec 50 % v/v de culture et 50% v/v de glycerol. Ces stocks on été conservés à -80°C dans des tubes à bouchon vissable dédiés.

```
100/1                                   130/11                                  160/21                                  190/31
ATG CGG GGT TCT CAT CAT CAT CAT CAT CAT GGT ATG GCT AGC ATG ACT GGT GGA CAG CAA ATG GGT CGG GAT CTG TAC GAC GAT GAC GAT AAG GAT CCG CCG GCT GAA TTT ATG AGC AAG
M  R  G  S  H  H  H  H  H  H  G  M  A  S  M  T  G  G  Q  Q  M  G  R  D  L  Y  D  D  D  D  K  D  P  P  A  E  F  M  S  K

220/41                                  250/51                                  280/61                                  310/71
GGC GAG GAG CTG TTC ACC GGG GTG GTG CCC ATC CTG GTC GAG CTG GAC GGC GAC GTA AAC GGC CAC AAG TTC AGC GTG TCC GGC GAG GGC GAG GGC GAT GCC ACC TAC GGC AAG CTG ACC
G  E  E  L  F  T  G  V  V  P  I  L  V  E  L  D  G  D  V  N  G  H  K  F  S  V  S  G  E  G  E  G  D  A  T  Y  G  K  L  T

340/81                                  370/91                                  400/101                                 430/111
CTG AAG TTC ATC TGC ACC ACC GGC AAG CTG CCC GTG CCC TGG CCC ACC CTC GTG ACC ACC CTG ACC TAC GGC GTG CAG TGC TTC AGC CGC TAC CCC GAC CAC ATG AAG CAG CAC GAC TTC
L  K  F  I  C  T  T  G  K  L  P  V  P  W  P  T  L  V  T  T  L  T  Y  G  V  Q  C  F  S  R  Y  P  D  H  M  K  Q  H  D  F

460/121                                 490/131                                 520/141                                 550/151
TTC AAG TCC GCC ATG CCC GAA GGC TAC GTC CAG GAG CGC ACC ATC TTC TTC AAG GAC GAC GGC AAC TAC AAG ACC CGC GCC GAG GTG AAG TTC GAG GGC GAC ACC CTG GTG AAC CGC ATC
F  K  S  A  M  P  E  G  Y  V  Q  E  R  T  I  F  F  K  D  D  G  N  Y  K  T  R  A  E  V  K  F  E  G  D  T  L  V  N  R  I

580/161                                 610/171                                 640/181                                 670/191
GAG CTG AAG GGC ATC GAC TTC AAG GAG GAC GGC AAC ATC CTG GGG CAC AAG CTG GAG TAC AAC TAC AAC AGC CAC AAC GTC TAT ATC ATG GCC GAC AAG CAG AAG AAC GGC ATC AAG GTG
E  L  K  G  I  D  F  K  E  D  G  N  I  L  G  H  K  L  E  Y  N  Y  N  S  H  N  V  Y  I  M  A  D  K  Q  K  N  G  I  K  V

700/201                                 730/211                                 760/221                                 790/231
AAC TTC AAG ATC CGC CAC AAC ATC GAG GAC GGC AGC GTG CAG CTC GCC GAC CAC TAC CAG CAG AAC ACC CCC ATC GGC GAC GGC CCC GTG CTG CTG CCC GAC AAC CAC TAC CTG AGC ACC
N  F  K  I  R  H  N  I  E  D  G  S  V  Q  L  A  D  H  Y  Q  Q  N  T  P  I  G  D  G  P  V  L  L  P  D  N  H  Y  L  S  T

820/241                                 850/251                                 880/261                                 910/271
CAG TCC GCC CTG AGC AAA GAC CCC AAC GAG AAG CGC GAT CAC ATG GTC CTG CTG GAG TTC GTG ACC GCC GCC GGG ATC ACT CTC GGC ATG GAC GAG CTG TAC AAG TAA
Q  S  A  L  S  K  D  P  N  E  K  R  D  H  M  V  L  L  E  F  V  T  A  A  G  I  T  L  G  M  D  E  L  Y  K  *
```

ATG Etiquette poly-histidine en N-terminal de l'eGFP.
ATG Site de clivage phénylalanine pour l'endopeptidase α–chymotrypsine.

Figure 10.1 – Séquence d'ADN insérée dans le plasmide pRSET-B qui code la protéine EGFP, l'étiquette histidine et le site de clivage à l'α-chymotrypsine en N-terminal

Les **figures 10.2** et **10.3** représentent la séquence d'acide aminés qui résultent de la séquence d'ADN présentée précédemment, avant et après clivage.

```
        10          20          30          40          50          60
MRGSHHHHHH GMASMTGGQQ MGRDLYDDDD KDPPAEFMSK GEELFTGVVP ILVELDGDVN
        70          80          90         100         110         120
GHKLTVSGEG EGDATYGKLT LKFICTTGKL PVPWPTLVTT LTYGVQCFSR YPDHMKQHDF
       130         140         150         160         170         180
FKSAMPEGYV QERTIFFKDD GNYKTRAEVK FEGDTLVNRI ELKGIDFKED GNILGHKLEY
       190         200         210         220         230         240
NYNSHNVYIM ADKQKNGIKV NFKIRHNIED GSVQLADHYQ QNTPIGDGPV LLPDNHYLST
       250         260         270
QSALSKDPNE KRDHMVLLEF VTAAGITLGM DELYK
```

Etiquette poly-histidine.
Site de clivage à l'α-chymotrypsine.
Mutations F64L, S65T et H231.

Figure 10.2 – Séquence d'acide aminés qui constitue la protéine EGFP avec son étiquette histidine et le site de clivage à l'α-chymotrypsine en N-terminal

```
        10          20          30          40          50          60
MSKGEELFTG VVPILVELDG DVNGHKFSVS GEGEGDATYG KLTLKFICTT GKLPVPWPTL
        70          80          90         100         110         120
VTTLTYGVQC FSRYPDHMKQ HDFFKSAMPE GYVQERTIFF KDDGNYKTRA EVKFEGDTLV
       130         140         150         160         170         180
NRIELKGIDF KEDGNILGHK LEYNYNSHNV YIMADKQKNG IKVNFKIRHN IEDGSVQLAD
       190         200         210         220         230
HYQQNTPIGD GPVLLPDNHY LSTQSALSKD PNEKRDHMVL LEFVTAAGIT LGMDELYK
```

Figure 10.3 – Séquence d'acide aminés qui constitue la protéine EGFP obtenue après clivage

Une expérience de biologie moléculaire à même été intentée afin d'obtenir une autre construction pour l'étiquette histidine en extrémité N-terminale de la protéine. Le but était de créer un site de clivage pour une enzyme bien plus spécifique que α-chymotrypsine (qui, pour rappel, coupe une chaîne polypeptidique après chaque acides aminés aromatiques accessibles). Mais par manque de temps, les premières expériences d'amplification d'ADN par PCR n'ayant pas donné des quantités d'ADN facilement exploitables, nous avons choisi de continuer les expériences avec la construction du Dr. Philippe Oger.

10.2.2 Production et purification de l'AvEGFP

Le protocole de production et de purification a été adapté de celui de R. M. Wachter mais celui de M. Ormö s'est aussi avéré parfaitement fonctionnel.

Toutes les valeurs de concentration en protéine qui vont suivre ont été calculées avec le coefficient d'extinction molaire théorique de l'EGFP à 488 nm, soit $\varepsilon_{488\,nm} = 63000\ M^{-1}.cm^{-1}$. Toutes les valeurs de densité optique ("DO") mentionnées ici ont été mesurée à $\lambda = 600$ nm.

Obtenir la protéine purifiée et concentrée par le protocole qui va suivre requiert trois jours pleins au minimum et se fait intégralement à 4°C. Il faut prévoir de réserver : un incubateur à 37°C pour une préculture sur la nuit (jour 0 au soir), un incubateur à 25°C pour la culture bactérienne (jour 1), une centrifugeuse HLA avec les tubes adéquats (jour 1), un micro-fluidizer ou un sonicateur pour la lyse bactérienne (jour 2), une ultracentrifugeuse avec les tubes adequats (jour 2), une colonne d'affinité au nickel et un appareil ÄKTA à 4°C (jour 2), une colonne de désalinisation (jour 2), une place sur une roue à 4°C pour l'étape de clivage (jour 2, soir et nuit), une colonne échangeuse de cations et un appareil ÄKTA à 4°C (jour 3), une centrifugeuse avec des tubes centricons (jour 3).

10.2.2.1 Description du milieu de culture et de l'ensemble des tampons utilisés

Le milieu de culture utilisé pour la production de la protéine est le milieu YT. Prévoir 4 litres de milieu pour la production et 500 mL pour la préculture, le tout nécessitant d'être autoclavé au préalable. Les **tableaux 10.1, 10.2 et 10.3** rendent compte de la composition de ce milieu de culture et de l'ensemble des tampons requis. Attention, pour le tampon de lyse, le PMSF doit être ajouté au dernier moment car bien qu'il inhibe un certain nombre de sérine protéases, il est rapidement dégradé. En plus du PMSF, j'ai toujours utilise un cachet du "Complete antiprotease cocktail" de la marque Roche™.

Milieu YT (pour 1 L)	**Tampon de lyse**
8 g Tryptone	50 mM HEPES pH 7,9
5 g Extrait de levure	300 mM NaCl
5 g NaCl	5 mM β-mercaptoethanol
Eau MilliQ q.s.p	0,1 mM PMSF
	Cocktail d'anti-protéases
	Eau MilliQ q.s.p

Tableau 10.1 – Composition du milieu de culture et du tampon de lyse

Tampon A HisTrap	**Tampon B HisTrap**	**Tampon de désalinisation**
50 mM HEPES pH 7,9	50 mM HEPES pH 7,9	50 mM HEPES pH 7,9
300 mM NaCl	300 mM NaCl	25 mM NaCl
5 mM β-mercaptoethanol	5 mM β-mercaptoethanol	Eau MilliQ q.s.p
Eau MilliQ q.s.p	250 mM Imidazole	
	Eau MilliQ q.s.p	

Tableau 10.2 – Composition des tampons pour la colonne HisTrap et la colonne de désalinisation

Tampon A Q-Sepharose	Tampon B Q-Sepharose	Tampon de stockage
50 mM HEPES pH 7,9	50 mM HEPES pH 7,9	20 mM HEPES pH 7,5
25 mM NaCl	500 mM NaCl	Eau MilliQ q.s.p
Eau MilliQ q.s.p	Eau MilliQ q.s.p	

Tableau 10.3 – Composition des tampons pour la colonne Q-Sepharose et du tampon de stockage de la protéine

10.2.2.2 Préculture sur la nuit

Prévoir un volume de supérieur à 120 mL de préculture pour une culture de 4 L à venir. Dans un erlenmeyer :

- ▷ plus de 120 mL de milieur de culture YT,
- ▷ 100 µg.mL^{-1} d'ampicilline,
- ▷ clone de EGFP ou aliquot de stock en glycerol,
- ▷ incubation sur la nuit à 37°C, 200 tpm.

10.2.2.3 Culture bactérienne et production de la protéine

Afin que l'oxygénation de la culture bactérienne soit optimale, des erlenmeyer de 5 L pour contenir 1 L de culture ont été systématiquement utilisés. Voici le protocole pour 1 L de culture :

- ▷ 1 L de milieu YT,
- ▷ 100 µg.mL^{-1} d'ampicilline,
- ▷ volume de préculture à ajouter pour que la densité optique théorique de départ pour chaque erlenmeyer soit de 0.1,
- ▷ incubation à 25°C, 200 tpm jusqu'à ce que DO $\approx$ 0,7 (3h40-4h),
- ▷ induction avec 1 mM d'IPTG,
- ▷ incubation à 25°C, 200 tpm, pendant au moins 5 h,
- ▷ centrifugation à 4°C, 5000 tpm pendant 30 min.

L'induction de 4 L de culture au moment où la DO atteint une valeur de 0,65 permet d'obtenir environ 8 g de culots bactériens vert fluo. Ces culots peuvent être conservés à -20°C dans des tubes Falcon® de 50 mL.

10.2.2.4 Lyse cellulaire

Les culots bactériens ont été resuspendus dans 50 mL du tampon de lyse. Durant tout le processus de lyse, j'ai toujours essayé de maintenir la suspension à basse température en utilisant de la glace.

Pour lyser les bacteries, un Micro-Fluidizer a été principalement utilisé mais, bien que moins efficace, la sonication conduit aussi à de bons résultats. Au Micro-Fluidizer, le même échantillon a subi deux passages sous une pression de 12000 psi. A la fin, l'échantillon doit être très liquide. Et afin de maintenir ce dernier le plus froid possible, j'ai toujours couvert le serpentin du Mirco-Fluidizer avec de la glace. Faire attention cependant à la dilution due au volume mort du Micro-Fluidizer. Dans le cas où un sonicateur a été utilisé, l'échantillon toujours la glace, j'ai toujours utilisé le programme suivant deux fois : une session de pulses de 2 min à 60 W pour 10 min de repos.

L'échantillon a ensuite été centrifugé à 4°C, 48400 tpm pendant 45 minutes et le surnageant résultant récupéré sans filtration (le culot ne se décollant pas du tube). Aussi, un aliquot de ce surnageant a été prélevé pour un futur dépôt sur gel de polyacrylamide 12 %.

10.2.2.5 Chromatographie sur colonne d'affinité au nickel

Afin de d'isoler la protéine EGFP grâce à son étiquette histidine, une colonne avec 10 mL de résine au nickel Ni-NTA Superflow® de marque Qiagen™ à été utilisée. Toute les étapes se font en chambre froide, à 4°C, sur un appareil ÄKTA délivrant un flux de 2 mL.min^{-1}.

La colonne est d'abord équilibrée avec 50 mL (cinq fois le volume de la colonne) de tampon A HisTrap. Le surnageant est alors passé sur la résine et le "passe-tout-droit" repassé encore une fois.

En second lieu, la colonne est lavée avec cette fois 100 mL de tampon A HisTrap.

Le gradient d'imidazole de 0 à 250 mM (tampon B HisTrap) est ensuite passé sur la colonne selon un flux de 2 mL.min^{-1} tout en éluant la protéine dans des fractions de 2 mL. Comme le montre la **figure 10.4**, l'élution de la protéine se fait en deux temps, entre 13 % (c.a.d 33 mM d'imidazole) et 86 % d'imidazole. Deux pics ont donc toujours été visibles sur les graphiques d'élution. Comme les fractions mélangées de ces deux pics m'ont permis d'obtenir des cristaux de EGFP suite à la première expérience de purification, je n'ai jamais pris le soin de séparer les fractions des deux pics. Mais il semble donc qu'il y ait deux populations différentes de EGFP.

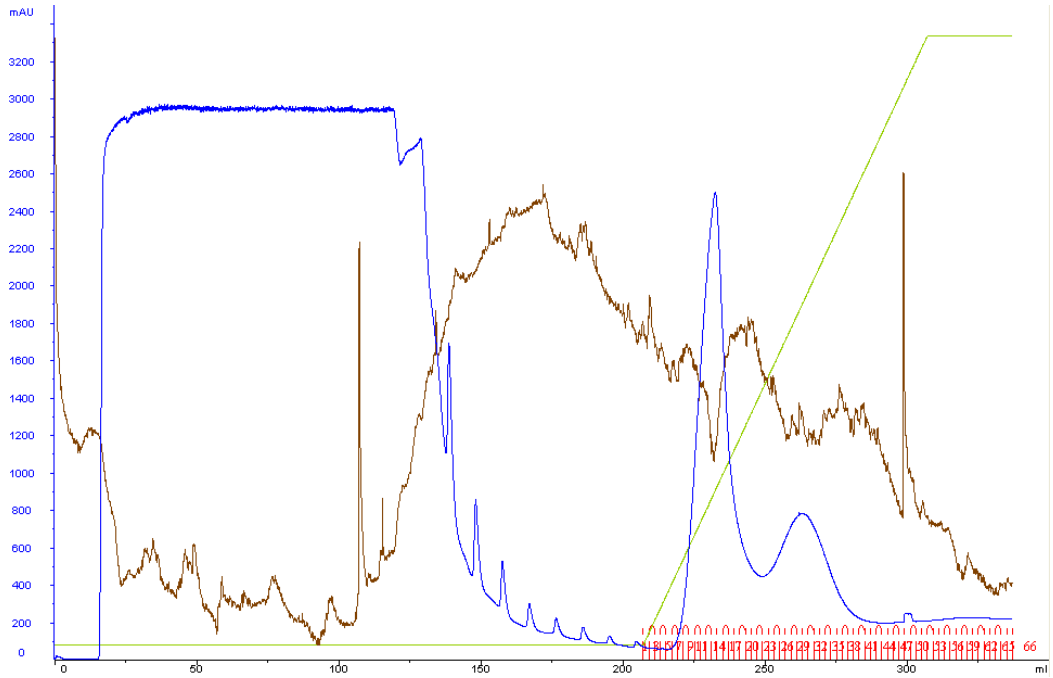


Figure 10.4 – Chromatogramme d'élution de l'EGFP à partir de la colonne d'affinité au nickel. Courbe bleue, suivi de la quantité en protéine à $\lambda_{UV} = 280$ nm. Courbe marron, mesure de la conductivité. Courbe verte, gradient d'imidazole.

Lors de la première expérience de purification, la quantité de protéine fut telle que le détecteur a été saturé lors du lavage de la colonne au tampon A HisTrap. J'ai donc du répéter toutes les étapes du protocole afin de ne rien gaspiller. Il est aussi à noter que purifier la protéine EGFP est simple et ludique. En effet, d'une part, le passage de la protéine est visible à lumière ambiante dans la colonne

et les tubulures et, d'autre part, si on dispose d'un appareil ÄKTA muni d'un détecteur à plusieurs longueurs d'onde, on peut en plus suivre spécifiquement la concentration en EGFP en utilisant aussi la détection à $\lambda = 488$ nm.

Des aliquots du "passe-tout-droit', du lavage au tampon A HisTrap et du mélange final des fractions ont toujours été prélevés pour analyse sur gel de polyacrylamide 12 %.

Pour 4 L de culture et 8 g de culots bactériens, entre 70 et 90 mL de EGFP étiquetée peuvent être récupérés par ce protocole.

10.2.2.6 Colonne de désalinisation

Afin de retirer un maximum d'imidazole de la solution riche en EGFP étiquetée, nocif pour la protéine à moyen terme, et afin de changer de tampon ladite solution, j'ai utilisé pour cette étape transitoire une colonne de désalinisation PD-10® (Amersham Biosciences™) à 4°C. Le protocole fourni avec la colonne PD-10® a été scrupuleusement suivi afin d'échanger le tampon. J'ai cependant du concentrer la solution de EGFP étiquetée au préalable afin d'obtenir l'échantillon de 2,5 mL requis. Pour cela, la protéine a été concentrée sur membrane tangentielle (tube centricon Millipore™ Amicon Ultra® de seuil de coupure de 10000 Da et de 15 mL) à 4°C et à 5000 tpm.

La colonne a été équilibrée avec 25 mL du tampon de désalinisation puis l'échantillon de 2,5 mL contenant la protéine étiquetée a été passé. Un aliquot de la solution résultante a été prélevée pour analyse sur un futur gel de polyacrylamide 12 %.

A la fin de cette étape, pour 4 L de culture initiaux, j'ai pu obtenir environ 10 mL de solution de EGFP étiquetée de 3,3 à 4,5 $mg.mL^{-1}$, soit 33 à 45 mg de protéine étiquetée.

10.2.2.7 Clivage de l'étiquette histidine

A la solution de EGFP étiquetée dépourvue d'imidazole a été ajouté, dans un tube Falcon® de 15 mL, l'endopeptidase α-chymotrypsine selon le ration 1 :50 w/w (1 mg d'enzyme pour 50 mg de EGFP de protéine). La poudre alors utilisée provient des laboratoires Sigma-Aldrich™. Le tube a été ensuite placé sur une roue, toujours en chambre froide, pendant 16 heures minimum.

Pour arrêter définitivement le clivage, 0,5 mM de PMSF ont été ajoutés. Attention cependant : lors de cette étape de clivage, bien utiliser le coefficient d'extinction molaire théorique à 488 nm ($\varepsilon_{488\,nm} = 63000\ M^{-1}.cm^{-1}$). Un aliquot de la solution résultante a systématiquement été prélevé.

10.2.2.8 Chromatographie sur colonne échangeuse de cations

Afin d'obtenir la protéine EGFP sans étiquette purifiée, nous avons utilisé une colonne Q-Sepharose échangeuse de cations. La colonne a été placée dans un appareil ÄKTA à 4°C avec le programme suivant : flux de 2 $mL.min^{-1}$, gradient complet entre le tampon A Q-Sepharose et le tampon B Q-Sepharose et élution de la protéine dans des fractions de 3 mL. Encore une fois, deux pics ont été visibles sur le chromatogramme, ce qui confirme la présence de deux population de EGFP. Mais ne voyant aucune différence sur un gel de polyacrylamide 12 %, l'ensemble des fractions correspondant aux deux pics ont été mélangées. Un aliquot de ce mélange à systématiquement été prélevé.

Cependant, afin de mieux séparer ces pics lors de cette étape, nous avons essayé d'étirer le gradient de sel (par paliers successifs) ou de le modifier mais je n'ai jamais réussi à avoir ces deux pics séparés.

Nous avons aussi appliqué le protocole de M. Örmo qui précède l'étape de purification sur colonne Q-Sepharose : faire passer la solution résultante du clivage une nouvelle fois sur colonne HisTrap. Les étiquettes histidines se retrouvent alors fixées à la résine alors que la protéine clivée se retrouve dans le passe-tout-droit. Pour ce faire, mon protocole a été le suivant : mélanger la solution résultante du clivage avec de la résine Ni-NTA dans un tube Falcon® de 15 mL placé sur une roue à 4°C. Après deux heures, ce tube a été centrifugé à 4°C pendant 30 minutes à 5000 tpm afin de ne prélever que le surnageant supposé contenir la protéine clivée. Malheureusement, cette étape supplémentaire ne m'a jamais permis d'obtenir de meilleurs résultats.

10.2.2.9 Concentration de la protéine EGFP purifiée

Afin de concentrer la protéine pour son utilisation dans de futures expériences de cristallisation, l'éluat de la chromatographie sur colonne Q-Sepharose a été placé dans un tube centricon Millipore™ Amicon Ultra® de seuil de coupure de 10000 Da et centrifugé à 4°C et à 5000 tpm. Cette étape permet aussi d'échanger le tampon pour le tampon de stockage final de la protéine. Un aliquot de cette solution a toujours été prélevé pour analyse sur gel de polyacrylamide 12 %.

En fin de purification, il m'a été possible d'obtenir 30 à 40 mg de EGFP purifiée pour 4 L de culture initiaux. La **figure 10.5** rend compte de la qualité de la pureté de la protéine.

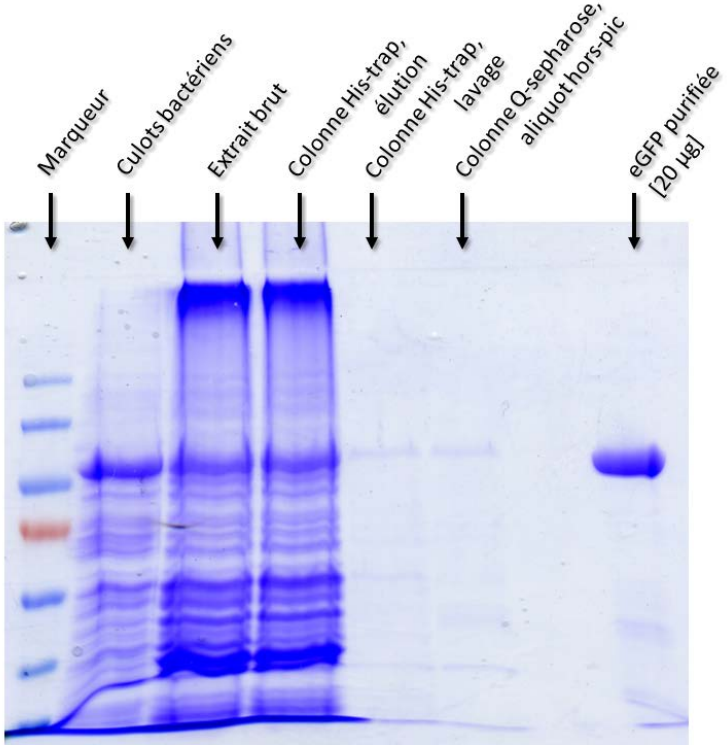


Figure 10.5 – Analyse de la pureté de la solution finale concentrée en EGFP sur gel de polyacrylamide 12 %

10.3 Forme soluble de la métalloprotéine membranaire CnrX du cpériplasme de la bactérie *Cupriavidus metallidurans* souche CH34 "CmCH34CnrXs"

La protéine CnrXs correspond à la partie cytoplasmique (soluble) de la métalloprotéine membranaire CnrX du périplasme de la bactérie *Cupriavidus metallidurans*. Elle est impliquée dans la régulation de l'expression de l'opéron cnrCBA qui code une pompe à efflux de cations qui participe à la résistance de la bactérie au cobalt et au nickel. La protéine, ancrée dans le périplasme sous forme homodimèrique, agit en tant que capteur d'ions métalliques dans le complexe qui initie la transduction du signal (Pompidor, 2008). Il s'agit d'une protéine de 33,2 kDa (2x16,6 kDa) dont le point isoélectrique théorique est de 7,04 et le coefficient d'extinction molaire à 280 nm $\varepsilon_{280\,nm}$ est de 15470 $M^{-1}.cm^{-1}$.

10.4 L'Acriflavine resistance protein B "EcAcrB"

EcAcrB est une protéine homotrimérique ancrée dans la membrane interne (cytoplasmique) de la bactérie à Gram négatif *Escherichia Coli.* L'analyse de sa séquence primaire révèle un poids moléculaire de 341 kDa (3147 acides aminés, 1049 par monomère), un point isoélectrique théorique de 5,39 et un coefficient d'extinction molaire mesuré est de 93000 $M^{-1}.cm^{-1}$.

Dans cette annexe sont consignés les protocoles mis au point pour obtenir de multiples cristaux de protéine EcAcrB parfaitement réguliers, pouvant avoisiner les 500 µm de longueur d'arête. Toutes les étapes qui vont suivre sont initialement inspirées du protocole de purification établi par Klaas M. Pos (Pos and Diederichs, 2002). Le protocole a été amélioré par Céline Julian-Binard et le Dr. Deniaud puis ensuite, par moi-même.

10.4.1 Production d'EcAcrB dans *E.coli*

J'ai utilisé la bactérie *E.coli*, souche CD43, prélevée à partir de solution-stock glycérol 50 % et transformée avec le vecteur d'expression modifié de Klaas M. Pos, le plasmide pET24. Ce plasmide porte trois cassettes : une cassette qui code la protéine EcAcrB avec 4 histidines en plus à l'extremité C-terminale, une cassette conférant la résistance à l'antibiotique kanamycine et un opéron, en 5' des deux autres, qui le rend fortement inductible à l'IPTG. Cette construction plasmidique permet donc de produire abondamment une protéine $EcAcrB_{His}$ avec une étiquette d'histidines qui permettra de la retenir par chromatographie d'affinité au nickel et de restreindre les cultures de bactéries à celles qui n'expriment que cette construction (celles pouvant se multiplier en présence de kanamycine).

La production de culots de bactéries exprimant la protéine $EcAcrB_{His}$ commence par une préculture de 100 mL en milieu TB (*"Terrific Broth", pour 1 L de milieu : 47,6 g de TB en poudre, 8 mL de glycérol 100 %, eau MilliQ q.s.p et autoclavage*) avec 50 $\mu g.mL^{-1}$ de kanamycine. La préculture est mise en incubateur sur la nuit, à 37°C, à 200 t.p.m.

On mesure la densité optique ("DO") de la préculture à la longueur d'onde $\lambda = 600$ nm à l'aide d'un spectrophotomètre et 3 L de cultures sont préparés avec 50 $\mu g.mL^{-1}$ de kanamycine. Un volume de préculture est ajouté dans chaque erlen de culture, volume calculé pour avoir une DO = 0,1 au départ de la culture. L'ensemble est mis en incubation à 37°C, 200 t.p.m. Après 3h15 d'incubation environ, la mesure de la DO indique une valeur d'environ 0,78. Les 3 L de culture sont alors placés en chambre

froide à 4°C et l'induction est réalisée avec 0,5 M d'IPTG. Une fois la température du milieu descendue à 30°C, la culture est mise à nouveau en incubateur, à 30°C pendant 3h30 et à 180 t.p.m, temps nécessaire à l'induction.

Tout le volume de culture est ensuite centrifugé à 6000 t.p.m, à 4°C, pendant 20 minutes. Les culots bactériens sont resuspendus dans un tampon A qui contient *20 mM Tris HCl pH 8,0 et 0,5 M NaCl*. Le choc osmotique alors produit permet de commencer la lyse cellulaire. Cette suspension est enfin centrifugée, à 4000 t.p.m, 4°C pendant 30 à 40 minutes.

Pour 3 L de culture, on obtient 12 à 16 g de culots bactériens.

10.4.2 Purification d'EcAcrB

La purification de la protéine membranaire EcAcrB est aisée, en quatre étapes sans recours à des systèmes automatisés de type HPLC ou FPLC. Les quatre étapes sont :

- ▷ La préparation et le dosage des fractions membranaires
- ▷ La solubilisation de ces dernières
- ▷ La purification de la protéine sur colonne d'affinité en résine de nickel ("HisTrap®")
- ▷ La concentration de la solution de protéine sur membrane tangentielle

10.4.2.1 Préparation et dosage des fractions membranaires de *E.coli*

Les culots bactériens sont mis en suspension dans un tampon B (*20 mM Tris HCl pH 8,0, 0,5 M NaCl, 2 mM $MgCl_2$, 3 comprimés de cocktail de protéases "Complete™" sans EDTA des laboratoires Roche®, traces de DNaseI*). La lyse des bactéries s'effectue en passant la solution en presse de French (11000 Psi) jusqu'à l'obtention d'une solution relativement limpide (deux passages sont généralement requis).

Une première centrifugation (10 minutes, 9000 t.p.m, 4°C) permet de séparer les bactéries non-lysées et les gros débris membranaires des fractions membranaires d'intérêt.

Le surnageant est alors soumis à une ultracentrifugation (1 heure, 29300 t.p.m, 4°C) dans le but de faire sédimenter les fractions membranaires. Les culots sont de couleur noir-orangée (l'intensité de la couleur orange peut servir pour juger de la quantité de protéines membranaires). Le surnageant est jeté.

Les culots de fractions membranaires sont resuspendus à leur tour dans 2 mL d'un tampon C (*20 mM Tris HCl pH 8,0 et 0,5 M NaCl*). La solution résultante est homogénéisée lentement au "Dounce". Cela permet de broyer de façon précise un maximum de fractions membranaires.

Enfin, le dosage des protéines dans les fractions membranaires se fait à l'aide du kit BCA (Sigma-Aldrich®). Cette méthode d'analyse spectroscopique s'appuie sur l'utilisation conjointe de l'acide bicinchoninique (acide 2,2'-biquinoline-4,4'-dicarboxylique ou "BCA") et du sulfate de cuivre pour suivre la réduction des ions Cu^{2+} en Cu^{1+} complexés aux protéines à pH basique. L'acide bicinchoninique forme un complexe bleu-violet dans le visible avec le Cu^{1+} et à pH basique. La mesure de l'intensité de la couleur relative à la quantité d'ions Cu^{2+} réduits, elle-même proportionnelle à la quantité de protéine, permet de mesurer la concentration en protéines en solution. Une gamme étalon réalisée avec quatre concentrations de Sérum Albumine Bovin (0-0,5-1-2 $mg.mL^{-1}$) auxquelles est ajouté le mélange BCA-$CuSO_4$ permet de calculer la concentration en fractions membranaires finales **[Figure 10.6]**.

Gamme BSA	
[BSA] mg/mL	Abs (562nm)
0	0
0,5	0,44
1	0,88
2	1,52

Echantillon	0,6

Soit: 0,73 mg/mL
Non dilué: 36,48 mg/mL

$y = 0{,}7566x + 0{,}048$
$x = (y - 0{,}048) / 0{,}7566$

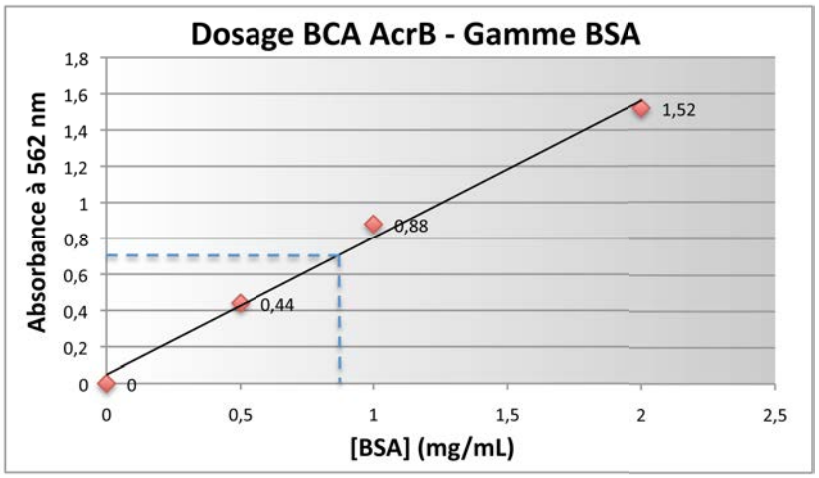


Figure 10.6 – Dosage BCA des fractions membranaires de *E.Coli*

Les solutions de la gamme étalon et notre échantillon sont mis à 37°C pendant 30 minutes. La mesure de l'absorbance de la solution se fait à la longueur d'onde $\lambda = 562$ nm. Dans l'exemple ci-dessus, 12,6 grammes de culots bactériens ont permis d'extraire 190 mg de fractions membranaires. Pour la suite, le rendement étant élevé, il est possible de ne prendre qu'un aliquot de 3 mL de fractions membranaires pour avoir un quantité suffisante de protéine EcAcrB en fin de purification. Ainsi, nous avons dilué la solution de fractions membranaires afin d'obtenir une solution de 6,33 mL à 30 mg.mL^{-1}. Le tampon C a été utilisé pour cette dilution. La solution peut être conservée à -80°C.

10.4.2.2 Solubilisation des fractions de membranes cytoplasmiques

Pour solubiliser les protéines membranaires contenues dans les fractions de membrane interne, nous utilisons le détergent ß-DDM. Le détergent n-Dodécyl-ß-maltoside a un poids moléculaire de 510 g.mol^{-1} et une Concentration Micellaire Critique ("CMC", concentration à partir de laquelle apparaissent les micelles de détergent) de 0,17 mM (0,0087 %) dans l'eau et de 0,12 mM (0,0062 %) dans une solution de 0,1 M NaCl (Zulauf, 1991). Ce détergent est particulièrement utilisé pour la solubilisation des protéines membranaires de la membrane interne des bactéries à Gram négatif. En formant des micelles dans lesquelles les parties transmembranaires des protéines viennent se loger, les détergents créent une "bouée" qui évitent l'agrégation du fait de ces parties hydrophobes (Pebay-Peyroula *et al.*, 1995; Penel *et al.*, 1998).

3 mL de la solution de fractions membranaires à 30 mg.mL^{-1} sont dilués dans 10 mL d'un tampon D (*50 mM Tris HCl pH 7,0, 10 mM imidazole pH 7,0, 10 % glycérol, 2 % ß-DDM et un demi comprimé de cocktail de protéases "cOmplete™" sans EDTA des laboratoires Roche®*). La solution résultante est mise à 4°C, sur tourniquet, pendant 1 heure. Elle est ensuite soumise à une ultracentrifugation (1 heure, 29000 t.p.m, 4°C) qui permet de culoter les bouts de membranes et les protéines non solubilisées. On récupère donc le surnageant contenant les protéines membranaires en micelles de détergent.

10.4.2.3 Purification de la protéine sur colonne d'affinité en résine de nickel

Toute la purification se fait en chambre froide, à 4°C et le débit imposé en sorti de colonne est un goutte-à-goutte lent (1,5 mL.min^{-1} environ). Pour 50 à 150 mg de fractions membranaires, 2 mL de

résine de nickel suffisent. J'ai utilisé la résine "Ni-NTA Superfow" de la marque Qiagen®. Pour cette étape de chromatographie d'affinité, trois tampons sont nécessaires pour fixer la protéine à la résine puis pour l'éluer par gradient d'imidazole :

- ▷ Tampon E : *20 mM Tris HCl pH 7,1, 10 mM imidazole pH 7,0, 10 % glycérol, 0,2 % ß-DDM*,
- ▷ Tampon F : *50 mM Tris HCl pH 7,1, 50 mM imidazole pH 7,0, 10 % glycérol, 0,2 % ß-DDM*,
- ▷ Tampon G : *50 mM Tris HCl pH 7,1, 200 mM imidazole pH 7,0, 10 % glycérol, 0,2 % ß-DDM.*

La colonne est équilibrée avec cinq fois son volume en tampon E (10 mL). La solution de protéines membranaires solubilisées est passée sur la colonne de résine de nickel. Pour fixer un maximum de protéine $EcAcrB_{His}$, le non-retenu est passé une autre fois sur la colonne. Un premier lavage au tampon E (35 mL) permet d'éluer la majeure partie des protéines membranaires non-souhaitées. Un second lavage au tampon F, légèrement plus concentré en imidazole, permet d'éluer un peu plus le non-spécifique. Enfin, l'élution de la protéine $EcAcrB_{His}$ se fait avec les 200 mM d'imidazole contenus dans le tampon G.

Comme le montre la photo du gel dénaturant de polyacrylamide 8 % **[figure 10.7]**, ce protocole de purification suffit amplement pour obtenir la protéine $EcAcrB_{His}$ pure (la composition du gel ainsi que celle du colorant au bleu de Coomassie utilisées sont données en annexe). Aucun protocole de purification ne prévoit, à ce jour, le clivage de l'étiquette histidine car il n'entrave en rien la cristallisation de la protéine.

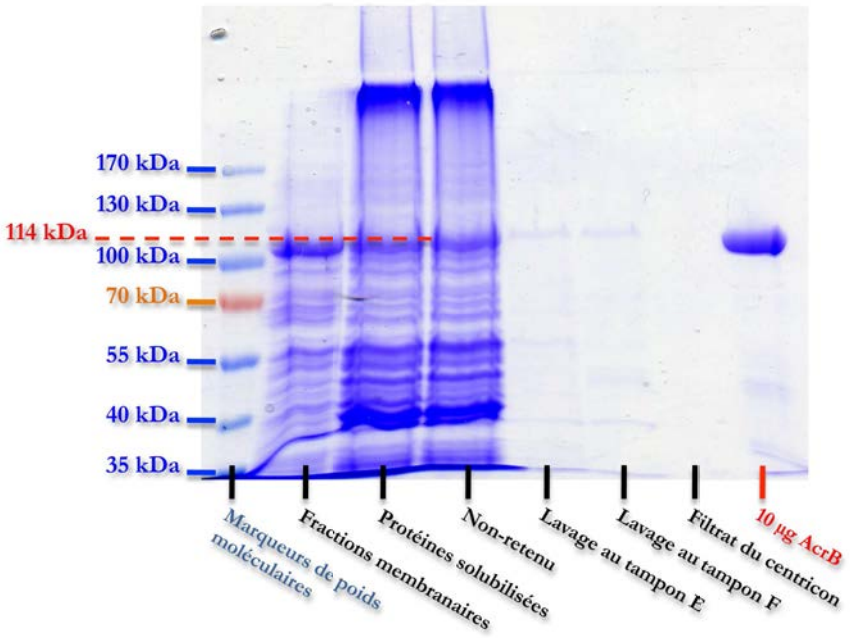


Figure 10.7 – Suivi de la purification de la protéine membranaire EcAcrB sur gel dénaturant de polyacrylamide 8 %

La dernière étape est la concentration de protéine $EcAcrB_{His}$ sur membrane tangentielle, étape qui permet aussi le changement de tampon afin de mettre la protéine dans son tampon de stockage. Dans un tube centricon de seuil de coupure 50 kDa, la solution finale est concentrée à 10 $mg.mL^{-1}$ dans

un tampon de stockage H contenant *10 mM HEPES pH 7,0 et 0,02 % ß-DDM*. Ainsi, pour 90 mg de fractions membranaires solubilisées, on obtient 2,05 mg de protéine membranaire EcAcrB$_{His}$ purifiée. En terme de volume, cela revient à obtenir 205 µL de solution de protéine EcAcrB$_{His}$ à 10 mg.mL^{-1} pour 3 mL de protéines membranaires solubilisées à 30 mg.mL^{-1}, ce qui équivaut à environ 1,5 L de culture bactérienne initiale.

10.5 Le lysozyme de blanc d'œuf de poule "HEWL"

Le lysozyme ("muramidase" ou encore " N-acetylmuramide glycanhydrolase") est une enzyme globulaire soluble eucaryote qui appartient à la famille des O-glycosyl hydrolases (classification enzymatique E.C. 3.2.1.17). Son poids moléculaire est de 14,3 kDa pour 129 acides aminés, avec un point isoélectrique théorique de 9.32 et un coefficient d'extinction molaire théorique à 280 nm $\varepsilon_{280\,nm}$ = 37970 M^{-1}.cm^{-1}. On la retrouve chez de nombreuses espèces et notamment en abondance dans les larmes, la salive, le lait maternel, les mucus et dans le blanc d'œuf (Lenka Grunclov *et al.*, 2003).

La première description de cette molécule revient à P. Laschtchenko en 1909 mais on doit l'attribution du nom "lysozyme" à A. Flemming qui, en 1922, observe qu'une entité présente dans les larmes et les sécrétions nasales provoque la lyse des bactéries (The Alexander Flemming papers, 2006). Première enzyme à avoir été totalement séquencée, il s'agit aussi de la deuxième structure résolue par diffraction des rayons X (à 2 Å de résolution) en 1965 par D. C. Phillips. Cette structure à permis de caractériser les acides aminés tryptophane 62 et 63 qui forment une véritable "pince à sucre" capable d'aller fixer le lysozyme sur les parois bactériennes pour hydrolyser la liaison 1-4 entre l'acide N-acetylmuramique et la partie N-acetyl-D-glucosamine des peptidoglycanes, provoquant ainsi la lyse de la bactérie.

10.6 La thaumatine de *Thaumatococcus daniellii* "TdTHAU"

La thaumatine est une protéine globulaire soluble eucaryote qui est extraite du cytoplasme des cellules qui constituent les graines du fruit de l'arbre *Thaumatococcus daniellii*, autrement appelé Katemfe. Composée de 207 acides aminés, son poids moléculaire est de 22,2 kDa, pour un point isoélectrique théorique de 8.5 et un coefficient d'extinction molaire à 280 nm $\varepsilon_{280\,nm}$ = 29420 M^{-1}.cm^{-1}. Elle est utilisée dans l'alimentation et la pharmacologie en tant qu'édulcorant, exhausteur de goût ou atténuateur d'amertume car il s'agit d'un édulcorant naturel ayant un pouvoir sucrant des plus élevés : entre deux mille et trois mille fois plus sucré que le saccharose à poids égal. Il s'agit en outre d'une protéine modèle pour la biocristallographie car les cristaux obtenus sont robustes et présentent un très beau profil de diffraction. La première structure publiée sur la Protein Data Bank a été résolue par diffraction des rayons X en 1989 et à 3,2 Å de résolution.

Chapitre 11

Exploration d'autres méthodes pour révéler la fixation des complexes de lanthanide avant l'étape de diffraction

Nous avons vu que la luminescence des complexes de la famille des picolinates d'europium ou de terbium permet de détecter leur fixation dans les cristaux dérivés de protéine. Cette propriété permet de savoir si le complexe a une affinité pour la protéine et permet de présumer un phasage expérimental ultérieur conduisant à la structure de celle-ci. Mais seule cette famille de complexe présente actuellement cette très intéressante propriété. Pour les autres complexes issus de l'imagerie médicale, il est nécessaire d'arriver jusqu'à la lourde étape de la diffraction pour savoir si le complexe se fixe à la protéine. Ainsi, la recherche de méthodes simples et peu coûteuses, permettant de savoir au préalable si un complexe de lanthanide peut se fixer à la protéine d'intérêt, a toujours été au centre de nos préoccupation et à déjà fait l'objet de la thèse de Meike Stelter. C'est donc en partie dans ce but qu'au cours de mon travail de recherche nous avons aussi réalisés des expériences de Résonnance Magnétique Nucléaire et de Résonnance Plasmonique de Surface avec les complexes de lanthanide.

Cette annexe contient donc une description très succincte de ces expériences et de l'ensemble des observations qui en découlent.

11.1 Détecter la fixation des complexes de lanthanide sur des macromolécules biologiques par Résonnance Plasmonique de Surface (Biacore®)

Je tiens à remercier ici le Dr. Eric Chabrol pour m'avoir non seulement chaperonné tout au long de ces expériences de Résonnance Plasmonique de Surface ("SPR", Biacore®) mais aussi pour m'avoir transmis une bonne partie de son savoir. Je remercie aussi le Dr. Nicole Thielens et le Dr. Romain Vives pour l'accès à la plateforme Biacore® de l'IBS et pour leur précieux conseils. Enfin, je remercie de Dr. Ieva Sutkeviciute pour m'avoir donné la composition de son mélange personnel pour régénérer les puces Biacore®.

11.1.1 Principe théorique de la SPR et fonctionnement du Biacore®

Afin d'étudier l'interaction de deux entités dans un flux et de calculer une constante d'affinité (K_d) pour ladite interaction, on greffe une des deux entité sur une puce Biacore®, une protéine par exemple, et on fait passer dans un flux au-dessus de cette puce, une autre entité, **figure 11.1**. Il n'y a donc pratiquement pas de flux au niveau de ce qui est greffé sur la puce. Toutes les entités qui peuvent interagir avec ce qui est greffé se retrouvent liées via le phénomène de sédimentation. Le flux n'intervient donc pas dans la modification de ce qu'on mesure : l'angle de réfraction.

En fonction de ce qui interagit, l'angle de réfraction ϑ_i entre l'onde réfléchie et la puce va changer et c'est cette variation de ϑ_i que l'on mesure. D'un point de vue plus physique, une liaison entre l'entité greffée et celle qui sédimente va modifier la propagation de l'onde évanescente dans le nuage d'électron que constitue une fine couche de métal (de l'or dans le cas d'une puce Biacore®). Les photons de cette onde évanescente vont en effet entrer en résonance avec ce nuage électronique et engendrer un "plasmon de surface", **figure 11.1**. C'est le phénomène de "résonance de plasmon de surface" ou "résonance plasmonique de surface".

De façon plus générale, un plasmon de surface se défini par une onde se propageant dans une densité d'électrons libres situés à l'interface entre un milieu conducteur, comme le métal, et un milieu diélectrique, comme l'air ou le verre.

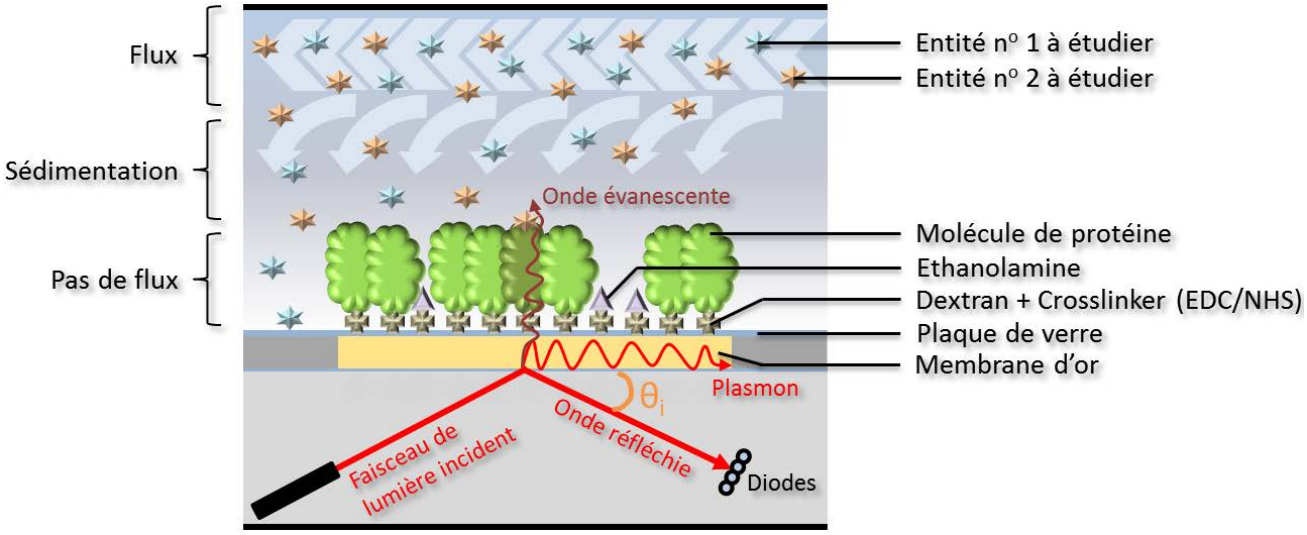


Figure 11.1 – Schéma du fonctionnement du Biacore®

L'apparition de ce plasmon de surface va modifier l'angle de réfraction ϑ_i et provoque une chute de l'intensité de l'onde réfléchie ce que l'on mesure grâce à une barrette de diodes, **figure 11.1 et 11.2**. La variation d'intensité dépend de ce qui se lie aux entités greffées (la taille, la forme et la quantité), du pH de la solution utilisée ou de la présence de certains ions (calcium...). Dans notre cas, cette donnée est à garder à l'esprit étant donné que nous utilisons des atomes de lanthanide.

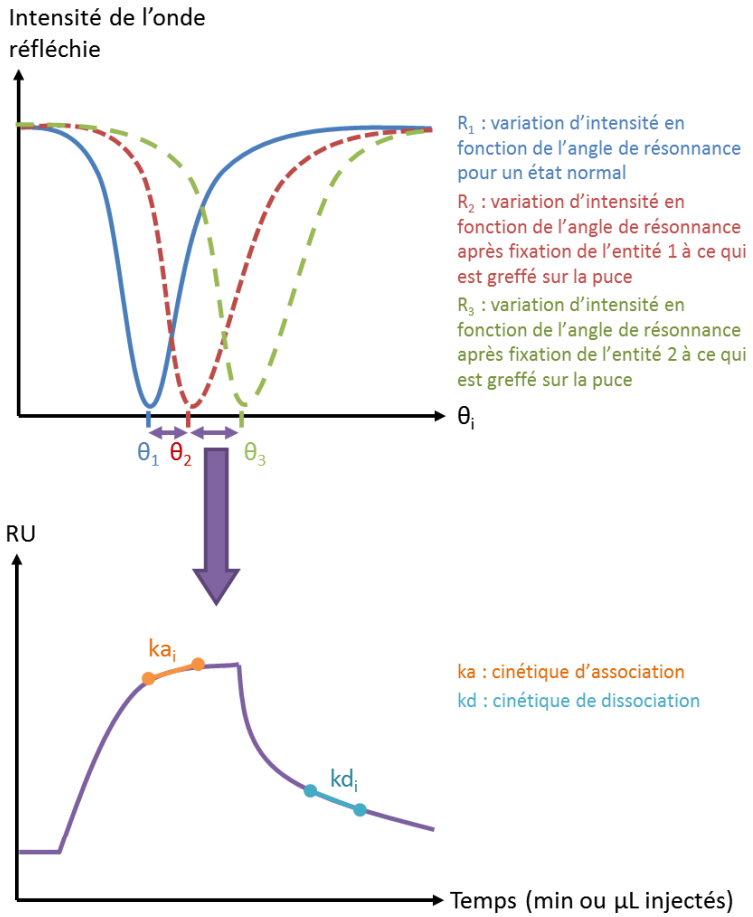

Intensité de l'onde réfléchie
R1 : variation d'intensité en fonction de l'angle de résonnance pour un état normal
R2 : variation d'intensité en fonction de l'angle de résonnance après fixation de l'entité 1 à ce qui est greffé sur la puce
R3 : variation d'intensité en fonction de l'angle de résonnance après fixation de l'entité 2 à ce qui est greffé sur la puce
θi
θ1 θ2 θ3
RU
kai
kdi
ka : cinétique d'association
kd : cinétique de dissociation
Temps (min ou µL injectés)

Figure 11.2 – Données mesurées pou pouvant être calculées par une expérience Biacore®

Aussi, la mesure du niveau d'interaction entre les deux entités en fonction du temps permet d'évaluer un taux d'association. Pour cela, tracer les valeurs que prend l'unité de résonance "RU" en fonction du temps permet de calculer une cinétique d'association "k_a" et une cinétique de dissociation "k_d", figure **11.2**. Une unité de résonance correspond à une variation de 0,0001° de la valeur de l'angle de réfraction. Connaissant la valeur de k_a et celle de k_d, on pourra alors calculer la valeur de la constante d'affinité (ou "constante de dissociation") K_d car :

$$K_d = \frac{k_d}{k_a}$$

Une entité qui se fixe rapidement et qui se dissocie tout aussi vite aura donc la même valeur de K_d qu'une entité qui s'associe lentement et se dissocie aussi lentement. Le calcul d'un K_d nécessite de faire une gamme de concentration pour l'entité injectée en flux et une certaine reproductibilité des résultats.

11.1.2 Protocole général

Le modèle d'appareil Biacore® qui a été utilisé est le Biacore® 3000, le plus sensible de la gamme en 2011. Tous les tampons et toutes les solutions qui ont été utilisés ont été systématiquement filtrées et dégazées au préalable. Nous avons utilisé une puce CM4 à 4 pistes, **figure 11.3**.

Ce type de puce coûte plus de 200 € et chaque piste peut recevoir le flux individuellement (1,2,3 ou 4) ou en série (1-2, 1-2-3 ou 1-2-3-4). Ces pistes peuvent être fonctionnalisées indépendamment (une

molécule différente peut être greffé par piste). Nous avons toujours utilisé la piste 1 comme piste témoin, c.a.d sans protéine.

Figure 11.3 – Schéma d'une puce Biacore® vue du dessus

Le protocole de fonctionnalisation qui a été utilisé est le suivant. Une protéine différente a été greffé sur chacune des piste sauf sur la piste 1 qui nous à servi de piste témoin. Au-dessus de la lame de verre, la puce est recouverte d'une matrice de dextran, polymère de dextrose qui possède un groupement carboxyle libre. Nous avons d'abord provoqué une réaction chimique visant à changer ce groupement carboxyle en ester en utilisant un agent de liaison ("crosslinker") : le mélange EDC/NHS (ethyl(dimethylaminopropyl) carbodiimide/N-Hydroxysuccinimide). Cet ester constitue un parfait groupe partant pour la création d'une liaison amide avec un groupement amine protoné disponible à la surface de la protéine. La greffe d'une protéine sur une puce Biacore® peut donc se schématisé ainsi : Puce-Dextran-EDC/NHS-amine de la protéine, **figure 11.1**. Afin que les groupements de la protéine soient protonés, nous avons toujours utilisé lors de cette étape un tampon pour la protéine dont le pH soit si possible à deux unités en dessous du point isoélectrique de la protéine. De plus, un surfactant (P20) a toujours été ajouté dans le tampon de greffe au dernier moment. Les surfactant permettent d'éviter toutes les interactions non-specifiques (hydrophobes) entre deux entités. Ainsi, en fixant au préalable ce surfactant P20 à la protéine, on s'assure qu'elle ne pourra pas se lier à une autre entité que la puce via ses groupements amine. En dernier lieu, une fois la protéine greffée, nous avons injecté de l'ethanolamine afin de saturer toutes les molécules de dextran esterifiées n'ayant pas fixé une molécule de protéine. Ainsi, sur la piste témoin (piste 1), seul l'ethanolamine est greffée à la puce.

11.1.3 Résumé très succinct des expériences menées

Le **tableau 11.4** résume toutes les interactions entre un complexe de lanthanide et une protéine que nous avons tenté de révéler par expériences au Biacore®. Après chaque passage d'un complexe de lanthanide et avant le passage d'un autre, la puce doit être régénérée, c.a.d nettoyée de toute molécule de complexe de lanthanide fixée ou non aux protéines. Je ne décrirai pas ici les nombreux protocoles de régénération de puce, d'essais de différents pH et de différentes concentrations en complexe car cela allongerait considérablement cette annexe.

	HEWL	**CaMalDH**	**HvMalDH**	**TdTHAU**	**AfUOx**
HMTDPA ligand seul	X				
$[Eu(HMTDPA)_3]^{3-}$	X				
$[Eu(DPA)_3]^{3-}$	X	X	X	X	X
GdDTPA-BMA	X	X	X		
GdDO3A	X	X	X		
GdHPDO3A	X	X	X	X	X

Figure 11.4 – Résumé des expériences menées au Biacore®. En colonne, les protéines modèles greffées sur la puce. En ligne, les entités passées en flux.

11.1.4 Résumé très succinct des résultats obtenus

Nous n'avons pu observer d'interaction entre le complexe GdDTPA-BMA et les protéines HEWL, CaMalDH et HvMalDH. Hors, grâce aux expériences de cristallographie menées au cours de cette thèse, nous savons que ce complexe se lie à la protéine HvMalDH.

Il est impossible de mesurer un quelconque signal avec le ligand HMTDPA seul, même sur la piste témoin (interaction avec l'ethanolamine). La taille du ligand doit être en deçà de la taille seuil d'une molécule permettant la mesure d'une interaction par expérience de Biacore®.

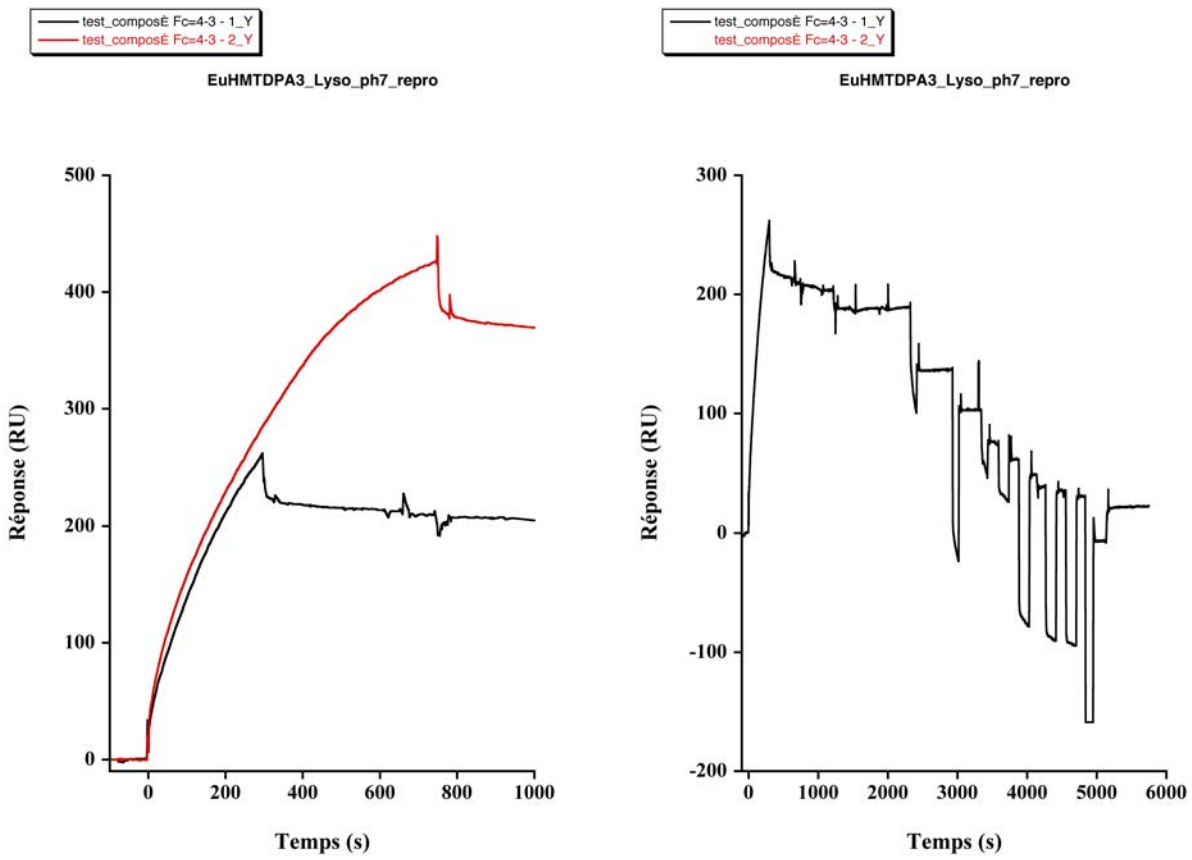
La protéine HEWL est un bon modèle car elle résiste aux multiples protocoles de régénération de puce.

Malgré les mauvais traitements du aux nombreuse tentative de régénération que la puce à subit, les expériences mettant en évidence l'interaction entre le complexe $[Eu(HMTDPA)_3]^{3-}$ et le HEWL sont reproductibles, **figures 11.5a et 11.5b**. En effet, à pH 4,0, après régénération, le complexe ajouté à une concentration doublement supérieure à la première conduit à l'observation d'un signal deux fois supérieur à celui observé avec la première concentration. Cependant, lors d'une même expérience à pH 7,5, ce facteur 2 entre les unités de résonance mesurée et la concentration n'est plus respecté. Il y a donc moins d'interaction entre le complexe et la protéine à pH 7,5 qu'à pH 4,0.

Bien qu'un signal ait pu être mesuré lors du passage du complexe $[Eu(HMTDPA)_3]^{3-}$ sur la piste contenant la protéine CaMalDH, il est impossible d'affirmer que le complexe se fixe. Compte tenu de l'extrême difficulté pour régénérer la puce, ce qui a tendance à dénaturer la protéine fixée, ce signal n'a pu être observé qu'une seul fois.

L'interaction entre le complexe GdDO3A et les protéines HEWL et CaMalDH a pu être mise en évidence par expérience de Biacore®. Le GdDO3A se fixe sur CaMalDH même à très faible concentration. Par contre, il y a un effet seuil avec le lysozyme : il faut charger l'injection avec au minimum 100 mM de complexe pour qu'il se fixe à la protéine.

L'interaction entre le complexe GdHPDO3A et les protéines HEWL et CaMalDH a aussi pu être mise en évidence par expérience de Biacore®. Même observation qu'avec le complexe GdDO3A, il y a un effet seuil avec le lysozyme. Il faut ajouter le complexe à plus de 100 mM pour détecter une interaction. Cependant, aucune interaction entre le GdHPDO3A et la protéine TdTHAU n'a pu être détectée.



(a) Interaction mesurée par Biacore® entre le complexe $[Eu(HMTDPA)_3]^{3-}$, injecté à deux concentrations différente et la protéine HEWL

(b) Régénération de la puce Biacore® où l'interaction entre le complexe $[Eu(HMTDPA)_3]^{3-}$ et la protéine HEWL a été observée

Figure 11.5 – Observation reproductible de l'intéraction entre le complexe $[Eu(HMTDPA)_3]^{3-}$ et la protéine HEWL par expérience de Biacore®

Avec le complexe $[Eu(DPA)_3]^{3-}$, il n'a pas été possible de mesuré une interaction avec les protéines et ce, quelle que soit la gamme de concentration en complexe. En effet, il y a eu davantage d'interaction entre le complexe et l'ethanolamine de la piste témoin qu'avec le lysozyme ou les autres protéines. De plus, que ce soit à pH 4,0 ou pH 7,0, le signal mesuré semble montrer une association et une dissociation trop rapide pour être censée.

11.1.5 Conclusion

De manière générale, voici ce qu'on peut retenir de ces expériences au Biacore®. Bien que les appareils aient évolués depuis les premiers essais effectués par Meike Stelter, il semble que la taille des complexes de lanthanide soit toujours limite pour détecter une interaction par une telle méthode. Il est donc possible de mesurer une interaction avec les complexes complets (ligand + ion lanthanide) mais il semble vain d'essayer d'observer une interaction avec le ligand seul. En outre, le signal mesuré est proportionnel à la masse et à la quantité de l'entité fixée à la puce. Ainsi, un problème relatif à la greffe de protéines sur la puce subsiste : la quantité parfois trop faible de protéine fixée sur la puce ne permet pas de mesurer un signal avec des complexes de faibles affinités. Enfin, un autre problème non-négligeable est la régénération de la puce Biacore® suite au passage des complexes de lanthanide. De

manière générale l'interaction entre les complexes et les protéines, lorsqu'elle a lieu, est telle qu'il devient difficile voir quasi-impossible de régénérer la puce. Cela signifie qu'il faudrait compter une puce CM4 pour déceler la fixation d'un seul complexe avant l'étape de diffraction. Hors, si la méthode Biacore® n'est pas si complexe à mettre en œuvre, elle devient dès lors extrêmement coûteuse.

11.2 Détecter la fixation des complexes de lanthanide sur des macromolécules biologiques par Résonance Magnétique Nucléaire

Nous avons mené des expériences de Résonance Magnétique Nucléaire ("RMN") préliminaires avec les complexes $[Eu(HMTDPA)_3]^{3-}$, $[Eu(DPA)_3]^{3-}$ et le HEWL. Le but principal de ces expériences était de mettre en évidence qu'en solution on constate aussi la différence, que nous avons révélé par cristallographie, entre l'interaction du complexe $[Eu(DPA)_3]^{3-}$avec le HEWL et celle du complexe $[Eu(HMTDPA)_3]^{3-}$avec le HEWL. Pour rappel, nous avons montré au cours de cette thèse qu'au niveau des même acides aminés tryptophane de la protéine HEWL, le complexe $[Eu(DPA)_3]^{3-}$se lie principalement par des interactions hydrogènes avec les acides aminés arginine et lysine voisins alors que le complexe $[Eu(HMTDPA)_3]^{3-}$se lie, non seulement par ce type d'interaction, mais aussi par des interactions de type π-π et CH-π.

Je tiens à remercie ici le Dr. Eric Condamine pour avoir si bien accueilli (avec dynamisme et bonne humeur) ce projet et pour avoir réalisé les expériences de RMN proprement-dites.

Toujours par soucis de ne pas alourdir davantage ce manuscrit, je ne décrirais cependant pas l'intégralité des expériences réalisées et seul un résultat préliminaire sera présenté à titre de preuve de concept.

11.2.1 Brève description du principe théorique

Nous avons utilisé ici une expérience de RMN bidimensionnelle : la séquence NOESY (Nuclear Overhauser Enhancement SpectroscopY). Elle permet de déterminer les corrélations par couplage spatial entre deux noyau électroniques espacés de moins de 5 Å. Cette méthode permet par exemple de distinguer deux hydrogènes proches dans l'espace (NH_βet NH_γ) de deux hydrogènes pouvant être séparés par de nombreuses liaisons mais proche dans l'espace par repliement de la molécule.

Dans le cadre de notre expérience, nous mesurons l'influence de la distance moyenne entre l'atome paramagnétique (l'atome de lanthanide) et son voisinage au cours du temps. Cette influence est visible sur un spectre RMN 2D sous forme de "tâche de corrélation" ou "raie". Comme les mesures se font au cours du temps par RMN, il est donc aussi possible d'avoir une idée de la dynamique d'une zone précise de la molécule.

11.2.2 Brève description du protocole retenu

Nos travaux ont d'abord consisté à reprendre ceux de Gottfried Otting sur l'étude de la fixation du complexe $[Y(DPA)_3]^{3-}$ sur le HEWL (Su XC *et al.*, 2009). Nous avons pour cela commencé par reproduire le protocole décrit pour nous placer ensuite dans les mêmes conditions. En effet, utiliser le même tampon et surtout le même pH que celui qui a permis de déterminer la structure de la protéine modèle auparavant est fondamental en RMN. Un changement de pH supérieur à 0,1 unité de pH nécessiterai de réattribuer

tous les acides aminés (soit de re-déterminer) de la protéine modèle, ce qui n'était pas notre but, la structure du lysozyme ayant déjà été déterminée par RMN (Su XC *et al.*, 2009).

En premier lieu, l'expérience NOESY est réalisée sur un mélange de HEWL avec le complexe $[Y(DPA)_3]^{3-}$. L'Yttrium est utilisé car, en tant qu'atome diamagnétique, il ne produit pas de déplacement chimique. Cela permet donc de mesurer seulement l'influence des ligands du complexe d'Yttrium sur les déplacements chimiques. Cela sert de témoin ou de "blanc". En second lieu, on reproduit la même expérience mais avec l'atome de lanthanide voulu à la place de l'Yttrium. Par comparaison, on voit l'exacte influence du complexe de lanthanide sur les atomes de la surface protéiques voisins.

Les expériences ont été réalisées à une fréquence d'irradiation de 600 MHz, à pH 7,2 et à 25°C.

11.2.3 Type de résultat observable par RMN

La **figure 11.6** présente le type de résultat que nous avons pu obtenir par cette méthode. Les conditions d'obtentions des spectres NOESY comparés ici sont décrites dans la partie supérieure de la figure.

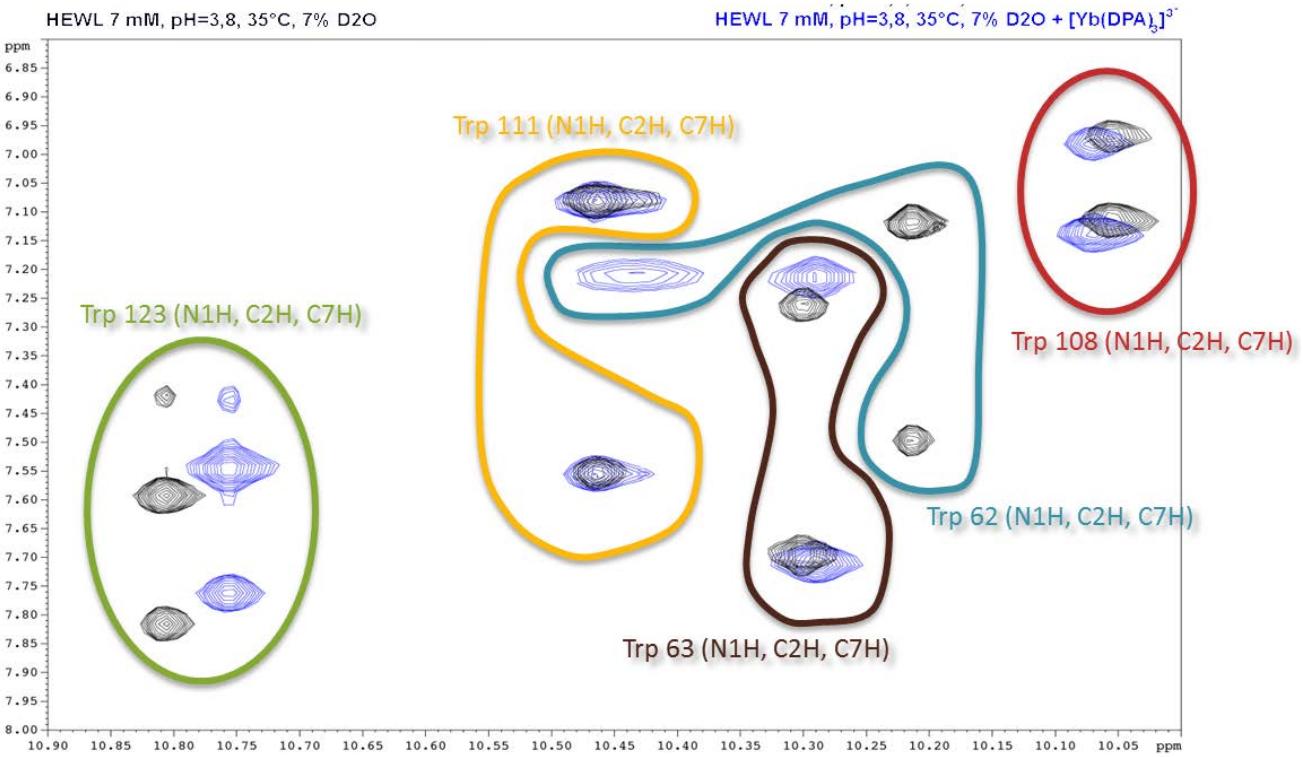

HEWL 7 mM, pH=3,8, 35°C, 7% D2O
HEWL 7 mM, pH=3,8, 35°C, 7% D2O + [Yb(DPA)3]3-
ppm
Trp 111 (N1H, C2H, C7H)
Trp 123 (N1H, C2H, C7H)
Trp 108 (N1H, C2H, C7H)
Trp 62 (N1H, C2H, C7H)
Trp 63 (N1H, C2H, C7H)

Figure 11.6 – Comparaison des tâche de corrélation correspondant aux atomes d'hydrogène des acides aminés tryptophane du HEWL déterminées lors de deux expérience NOESY, une en absence de complexe de lanthanide (en noir), l'autre en présence du complexe $[Yb(DPA)_3]^{3-}$ (en bleu).

On voit très nettement l'influence de la fixation du complexe $[Yb(DPA)_3]^{3-}$ à la surface du HEWL sur la position des tâches de corrélation correspondant aux atomes d'hydrogène des acides aminés tryptophane du site de fixation.

11.2.4 Conclusion

Par une expérience de résonance magnétique nucléaire bidimensionnelle de type NOESY, il semble donc tout à fait possible de révéler la fixation d'un complexe de lanthanide sur une protéine en solution. Cependant, il faut aussi garder à l'esprit qu'une expérience de RMN requiert une très grande quantité de protéine, quantité qu'il est rarement aisé d'obtenir.

Bibliographie

Abrahams, J. P. and Leslie, A. (**1996**) Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallographica Section D Biological Crystallography* **52** : 30–42.

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (**2010**) PHENIX : a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D Biological Crystallography* **66** : 213–221.

Allen, K. N. and Imperiali, B. (**2010**) Lanthanide-tagged proteins, an illuminating partnership. *Current Opinion in Chemical Biology* **14** : 247–254.

Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (**2001**) Electrostatics of nanosystems : application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences of the United States of America* **98** : 10037–10041.

Barthelmes, K., Reynolds, A. M., Peisach, E., Jonker, H. R. A., Denunzio, N. J., Allen, K. N., Imperiali, B., and Schwalbe, H. (**2010**) Engineering encodable lanthanide-binding tags into loop regions of proteins. *Journal of the American Chemical Society* **133** : 808–819.

Bianchi, A., Calabi, L., Corana, F., Fontana, S., Losi, P., Maiocchi, A., Paleari, L., and Valtancoli, B. (**2000**) Thermodynamic and structural properties of Gd(III) complexes with polyaminopolycarboxylic ligands : basic compounds for the development of MRI contrast agents. *Coordination Chemistry Reviews* **204** : 309–393.

Bijvoet, J. M., Peerdeman, A. F., and van Bommel, A. J. (**1951**) Determination of the absolute configuration of optically active compounds by means of X-rays. *Nature* **168** : 271–272.

Binnemans, K. and Görller-Walrand, C. (**1995**) On the color of the trivalent lanthanide ions. *Chemical physics letters* **235** : 163–174.

Blundell, T. L. and Johnson, L. N. (**1976**) Protein Crystallography. Academic Press.

Bonnete, F., Vivares, D., Robert, C., and Colloc'h, N. (**2001**) Interactions in solution and crystallization of Aspergillus flavus urate oxidase. *Journal Of Crystal Growth* **232** : 330–339.

Borissenko, L. and Groll, M. (**2005**) Crystal structure of TET protease reveals complementary protein degradation pathways in prokaryotes. *Journal of Molecular Biology* **346** : 1207–1219.

Bricogne, G. (**1988**) A Bayesian statistical theory of the phase problem. I. A multichannel maximum-entropy formalism for constructing generalized joint probability distributions of structure factors. *Acta crystallographica Section A, Foundations of crystallography* **44** : 517–545.

Brown, M., Peierls, R., and Stern, E. (**1977**) White lines in x-ray absorption. *Physical Review B* **15** : 738–744.

Cavicchioli, R. (**2007**) Archaea. molecular and cellular biology, Amer Society for Microbiology.

Chamas, Z. E. A., Guo, X., Canet, J.-L., Gautier, A., Boyer, D., and Mahiou, R. (**2010**) Clicked dipicolinic antennae for lanthanide luminescent probes. *Dalton Transactions* **39** : 7091–7097.

Charron, C., Giegé, R., and Lorber, B. (**2004**) Structure of thaumatin in a hexagonal space group : comparison of packing contacts in four crystal lattices. *Acta Crystallographica Section D Biological Crystallography* **60** : 83–89.

Collaborative Computational Project, Number 4 (**1994**) The CCP4 suite : programs for protein crystallography. *Acta Crystallographica Section D Biological Crystallography* **50** : 760–763.

Contreras-Martel, C., Job, V., Di Guilmi, A., Vernet, T., Dideberg, O., and Dessen, A. (**2006**) Crystal structure of penicillin-binding protein 1a (PBP1a) reveals a mutational hotspot implicated in beta-lactam resistance in Streptococcus pneumoniae. *Journal of Molecular Biology* **355** : 684–696.

Coquelle, N., Talon, R., Juers, D. H., Girard, E., Kahn, R., and Madern, D. (**2010**) Gradual adaptive changes of a protein facing high salt concentrations. *Journal of Molecular Biology* **404** : 493–505.

Costenaro, L., Zaccai, G., and Ebel, C. (**2002**) Link between protein-solvent and weak protein-protein interactions gives insight into halophilic adaptation. *Biochemistry* **41** : 13245–13252.

Cousins, K. R. (**2011**) Computer review of ChemDraw Ultra 12.0. *Journal of the American Chemical Society* **133** : 8388.

Cowtan, K. (**2006**) The Buccaneer software for automated model building. *Acta Crystallographica Section D Biological Crystallography* **62** : 1002–1011.

Cowtan, K. D. and Main, P. (**1996**) Phase combination and cross validation in iterated density-modification calculations. *Acta Crystallographica Section D Biological Crystallography* **52** : 43–48.

D'Aléo, A., Pompidor, G., Elena, B., Vicat, J., Baldeck, P. L., Toupet, L., Kahn, R., Andraud, C., and Maury, O. (**2007**) Two-photon microscopy and spectroscopy of lanthanide bioprobes. *ChemPhysChem* **8** : 2125–2132.

Dalhus, B., Saarinen, M., Sauer, U. H., Eklund, P., Johansson, K., Karlsson, A., Ramaswamy, S., Bjørk, A., Synstad, B., Naterstad, K., Sirevåg, R., and Eklund, H. (**2002**) Structural basis for thermophilic protein stability : structures of thermophilic and mesophilic malate dehydrogenases. *Journal of Molecular Biology* **318** : 707–721.

Dauter, Z. (**2002**) New approaches to high-throughput phasing. *Current opinion in structural biology* **12** : 674–678.

Dauter, Z., Dauter, M., and Dodson, E. (**2002**) Jolly SAD. *Acta Crystallographica Section D Biological Crystallography* **58** : 494–506.

Dauter, Z., Dauter, M., and Rajashankar, K. R. (**2000**) Novel approach to phasing proteins : derivatization by short cryo-soaking with halides. *Acta Crystallographica Section D Biological Crystallography* **56** : 232–237.

de La Fortelle, E. (**1997**) SHARP : A maximum-likelihood heavy-atom parameter refinement program for the MIR and MAD methods. *Methods in Enzymology* **276** : 472–494.

Delfosse, V., Girard, E., Birck, C., Delmarcelle, M., Delarue, M., Poch, O., Schultz, P., and Mayer, C. (**2009**) Structure of the archaeal pab87 peptidase reveals a novel self-compartmentalizing protease family. *PloS one* **4** : e4712.

Dolinsky, T. J., Nielsen, J. E., McCammon, J. A., and Baker, N. A. (**2004**) PDB2PQR : an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Research* **32** : 665–667.

Doublié, S. (**1997**) Preparation of selenomethionyl proteins for phase determination. *Methods in enzymology* **276** : 523–530.

Doutch, J., Hough, M. A., Hasnain, S. S., and Strange, R. W. (**2012**) Challenges of sulfur SAD phasing as a routine method in macromolecular crystallography. *Journal of synchrotron radiation* **19** : 19–29.

Durá, M. A., Receveur-Bréchot, V., Andrieu, J.-P., Ebel, C., Schoehn, G., Roussel, A., and Franzetti, B. (**2005**) Characterization of a TET-like aminopeptidase complex from the hyperthermophilic archaeon Pyrococcus horikoshii. *Biochemistry* **44** : 3477–3486.

Durá, M. A., Rosenbaum, E., Larabi, A., Gabel, F., Vellieux, F. M. D., and Franzetti, B. (**2009**) The structural and biochemical characterizations of a novel TET peptidase complex from Pyrococcus horikoshii reveal an integrated peptide degradation system in hyperthermophilic Archaea. *Molecular Microbiology* **72** : 26–40.

Elsliger, M. A., Wachter, R. M., Hanson, G. T., Kallio, K., and Remington, S. J. (**1999**) Structural and spectral response of green fluorescent protein variants to changes in pH. *Biochemistry* **38** : 5296–5301.

Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (**2010**) Features and development of Coot. *Acta Crystallographica Section D Biological Crystallography* **66** : 486–501.

Evans, G. and Pettifer, R. F. (**2001**) CHOOCH : a program for deriving anomalous-scattering factors from X-ray fluorescence spectra. *Journal Of Applied Crystallography* **34** : 82–86.

Evans, P. (**2005**) Scaling and assessment of data quality. *Acta Crystallographica Section D Biological Crystallography* **62** : 72–82.

Eventoff, W., Rossmann, M. G., Taylor, S. S., Torff, H. J., Meyer, H., Keil, W., and Kiltz, H. H. (**1977**) Structural adaptations of lactate dehydrogenase isozymes. *Proceedings of the National Academy of Sciences of the United States of America* **74** : 2677–2681.

Ferguson, A. D., Hofmann, E., Coulton, J. W., Diederichs, K., and Welte, W. (**1998**) Siderophore-mediated iron transport : crystal structure of FhuA with bound lipopolysaccharide. *Science* **282** : 2215–2220.

Feterman, G. and Hazan, M. (**2010**) La nature au fil des couleurs. Dakota Edition.

Fralick, J. A. (**1996**) Evidence that TolC is required for functioning of the Mar/AcrAB efflux pump of Escherichia coli. *Journal of Bacteriology* **178** : 5803–5805.

Franzetti, B., Schoehn, G., Hernandez, J.-F., Jaquinod, M., Ruigrok, R. W. H., and Zaccai, G. (**2002**) Tetrahedral aminopeptidase : a novel large protease complex from archaea. *The EMBO journal* **21** : 2132–2138.

Garman, E. and Murray, J. W. (**2003**) Heavy-atom derivatization. *Acta Crystallographica Section D Biological Crystallography* **59** : 1903–1913.

Garman, E. F. (**2010**) Radiation damage in macromolecular crystallography : what is it and why should we care ? *Acta Crystallographica Section D Biological Crystallography* **66** : 339–351.

Ginzburg, M. and Sachs, L. (**1970**) Ion Metabolism in a Halobacterium. I. Influence of age of culture on intracellular concentrations. *The Journal of General Physiology* **55** : 187–207.

Girard, E., Anelli, P. L., Vicat, J., and Kahn, R. (**2003**a) High-phasing-power lanthanide derivatives : taking advantage of ytterbium and lutetium for optimized anomalous diffraction experiments using synchrotron radiation. *Acta Crystallographica Section D Biological Crystallography* **59** : 1877–1880.

Girard, E., Stelter, M., Anelli, P. L., Vicat, J., and Kahn, R. (**2003**b) A new class of gadolinium complexes employed to obtain high-phasing-power heavy-atom derivatives : results from SAD experiments with hen egg-white lysozyme and urate oxidase from Aspergillus flavus. *Acta Crystallographica Section D Biological Crystallography* **59** : 118–126.

Glover, C. A. P., Postis, V. L. G., Charalambous, K., Tzokov, S. B., Booth, W. I., Deacon, S. E., Wallace, B. A., Baldwin, S. A., and Bullough, P. A. (**2011**) AcrB contamination in 2-D crystallization of membrane proteins : lessons from a sodium channel and a putative monovalent cation/proton antiporter. *Journal Of Structural Biology* **176** : 419–424.

Green, D. W., Ingram, V. M., and Perutz, M. F. (**1954**) The structure of Haemoglobin. IV. sign determination by the isomorphous replacement method. *Proceedings of the Royal Society A : Mathematical, Physical and Engineering Sciences* **225** : 287–307.

Grosse-Kunstleve, R. W. and Adams, P. D. (**2003**) Substructure search procedures for macromolecular structures. *Acta Crystallographica Section D Biological Crystallography* **59** : 1966–1973.

Grunclová, L., Fouquier, H., Hypsa, V., and Kopácek, P. (**2003**) Lysozyme from the gut of the soft tick Ornithodoros moubata : the sequence, phylogeny and post-feeding regulation. *Developmental and comparative immunology* **27** : 651–660.

Hendrickson, W. A. (**1979**) Phase information from anomalous-scattering measurements. *Acta Crystallographica Section A* **35** : 245–247.

Hendrickson, W. A. (**1981**) Structure of the hydrophobic protein crambin determined directly from the anomalous scattering of sulphur. *Nature* **290** : 107–113.

Hendrickson, W. A., Horton, J. R., and LeMaster, D. M. (**1990**) Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD) : a vehicle for direct determination of three-dimensional structure. *The EMBO journal* **9** : 1665–1672.

Hendrickson, W. A. and Lattman, E. E. (**1970**) Representation of phase probability distributions for simplified combination of independent phase information. *Acta Crystallographica Section B Structural Crystallography and Crystal Chemistry* **26** : 136–143.

Hoppe, W. (**1975**) Anomalous Scattering. Ramaseshan S. & Abrahams S. C.

Horrocks, W. D. (**1993**) Luminescence spectroscopy. *Methods in enzymology* **226** : 495–538.

Horrocks, W. D. and Sudnick, D. R. (**1981**) Lanthanide ion luminescence probes of the structure of biological macromolecules. *Accounts of Chemical Research* **14** : 384–392.

Jeudy, S., Stelter, M., Coutard, B., Kahn, R., and Abergel, C. (**2005**) Preliminary crystallographic analysis of the Escherichia coli YeaZ protein using the anomalous signal of a gadolinium derivative. *Acta crystallographica Section F, Structural biology and crystallization communications* **61** : 848–851.

Joyce, M. G., Radaev, S., and Sun, P. D. (**2010**) A rational approach to heavy-atom derivative screening. *Acta Crystallographica Section D Biological Crystallography* **66** : 358–365.

Kabsch, W. (**1993**) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *Journal Of Applied Crystallography* **26** : 795–800.

Kahn, R., Fourme, R., Bosshard, R., Chiadmi, M., Risler, J. L., Dideberg, O., and Wery, J. P. (**1985**) Crystal structure study of Opsanus tau parvalbumin by multiwavelength anomalous diffraction. *FEBS letters* **179** : 133–137.

Karle, J. (**1980**) Some developments in anomalous dispersion for the structural investigation of macromolecular systems in biology. *International Journal of Quantum Chemistry* **18** : 357–367.

Ko, T. P., Day, J., Greenwood, A., and McPherson, A. (**1994**) Structures of three crystal forms of the sweet protein thaumatin. *Acta Crystallographica Section D Biological Crystallography* **50** : 813–825.

Koch, M., Diez, J., Wagner, A., and Fritz, G. (**2010**) Crystallization and calcium/sulfur SAD phasing of the human EF-hand protein S100A2. *Acta crystallographica Section F, Structural biology and crystallization communications* **66** : 1032–1036.

Kramers, H. (**1927**) La diffusion de la lumiere par les atomes. *Atti Cong Intern Fisica* **2** : 545–557.

Kronig, R. d. L. (**1926**) On the Theory of Dispersion of X-Rays. *Journal of the Optical Society of America* **12** : 547–556.

Lakomek, K., Dickmanns, A., Mueller, U., Kollmann, K., Deuschl, F., Berndt, A., Lübke, T., and Ficner, R. (**2009**) De novo sulfur SAD phasing of the lysosomal 66.3 kDa protein from mouse. *Acta Crystallographica Section D Biological Crystallography* **65** : 220–228.

Larson, S. B., Day, J. S., Cudney, R., and McPherson, A. (**2007**) A novel strategy for the crystallization of proteins : X-ray diffraction validation. *Acta Crystallographica Section D Biological Crystallography* **63** : 310–318.

Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (**1993**) PROCHECK : a program to check the stereochemical quality of protein structures. *Journal Of Applied Crystallography* **26** : 283–291.

Lewis, W. G., Green, L. G., Grynszpan, F., Radić, Z., Carlier, P. R., Taylor, P., Finn, M. G., and Sharpless, K. B. (**2002**) Click chemistry in situ : acetylcholinesterase as a reaction vessel for the selective assembly of a femtomolar inhibitor from an array of building blocks. *Angewandte Chemie (International ed in English)* **41** : 1053–1057.

Lye, R. C., Phillips, J. C., Kaplan, D., Doniach, S., and Hodgson, K. O. (**1980**) White lines in L-edge x-ray absorption spectra and their implications for anomalous diffraction studies of biological materials. *Proceedings of the National Academy of Sciences of the United States of America* **77** : 5884–5888.

Madern, D. (**2002**) Molecular evolution within the L-malate and L-lactate dehydrogenase super-family. *Journal of molecular evolution* **54** : 825–840.

Matsuoka, D. and Nakasako, M. (**2009**) Probability distributions of hydration water molecules around polar protein atoms obtained by a database analysis. *The Journal of Physical Chemistry B* **113** : 11274–11292.

McPherson, A. and Cudney, B. (**2006**) Searching for silver bullets : an alternative strategy for crystallizing macromolecules. *Journal Of Structural Biology* **156** : 387–406.

Mie, G. (**1908**) Beiträge zur optik trüber medien, speziell kolloidaler metallösungen. *Annalen der Physik* **330** : 377–445.

Miller, R., Shah, N., Green, M. L., Furey, W., and Weeks, C. M. (**2007**) Shake-and-Bakeon the grid. *Journal Of Applied Crystallography* **40** : 938–944.

Molina, R., Stelter, M., Kahn, R., and Hermoso, J. A. (**2009**) Characterization of gadolinium complexes for SAD phasing in macromolecular crystallography : application to CbpF. *Acta Crystallographica Section D Biological Crystallography* **65** : 823–831.

Morales, R., Berna, A., Carpentier, P., Contreras-Martel, C., Renault, F., Nicodeme, M., Chesne-Seck, M.-L., Bernier, F., Dupuy, J., Schaeffer, C., Diemer, H., Van-Dorsselaer, A., Fontecilla-Camps, J. C., Masson, P., Rochu, D., and Chabriere, E. (**2006**) Serendipitous discovery and X-ray structure of a human phosphate binding apolipoprotein. *Structure* **14** : 601–609.

Morris, A. L., MacArthur, M. W., Hutchinson, E. G., and Thornton, J. M. (**1992**) Stereochemical quality of protein structure coordinates. *Proteins : Structure, Function, and Genetics* **12** : 345–364.

Murakami, S., Nakashima, R., Yamashita, E., Matsumoto, T., and Yamaguchi, A. (**2006**) Crystal structures of a multidrug transporter reveal a functionally rotating mechanism. *Nature* **443** : 173–179.

Murakami, S., Nakashima, R., Yamashita, E., and Yamaguchi, A. (**2002**) Crystal structure of bacterial multidrug efflux transporter AcrB. *Nature* **419** : 587–593.

Nagem, R. A. P., Dauter, Z., and Polikarpov, I. (**2001**) Protein crystal structure solution by fast incorporation of negatively and positively charged anomalous scatterers. *Acta Crystallographica Section D Biological Crystallography* **57** : 996–1002.

Nakasako, M. (**2004**) Water-protein interactions from high-resolution protein crystallography. *Philosophical transactions of the Royal Society of London Series B, Biological sciences* **359** : 1191–1204.

Nikaido, H. and Zgurskaya, H. I. (**2001**) AcrAB and related multidrug efflux pumps of Escherichia coli. *Journal of molecular microbiology and biotechnology* **3** : 215–218.

Okaya, Y. and Pepinsky, R. (**1956**) New formulation and solution of the phase problem in X-ray analysis of noncentric crystals containing anomalous scatterers. *Physical Review* **103** : 1645–1647.

Okusu, H. and Ma, D. (**1996**) AcrAB efflux pump plays a major role in the antibiotic resistance phenotype of Escherichia coli multiple-antibiotic-resistance (Mar) mutants. *Journal of Bacteriology* **178** : 306–308.

Ormö, M., Cubitt, A. B., Kallio, K., Gross, L. A., Tsien, R. Y., and Remington, S. J. (**1996**) Crystal structure of the Aequorea victoria green fluorescent protein. *Science* **273** : 1392–1395.

Painter, J. and Merritt, E. (**2006**a) TLSMD web server for the generation of multi-group TLS models. *Journal Of Applied Crystallography* **39** : 109–111.

Painter, J. and Merritt, E. A. (**2006**b) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallographica Section D Biological Crystallography* **62** : 439–450.

Patterson, A. (**1934**) A fourier series method for the determination of the components of interatomic distances in crystals. *Physical Review* **46** : 372–376.

Pebay-Peyroula, E., Garavito, R. M., Rosenbusch, J. P., Zulauf, M., and Timmins, P. A. (**1995**) Detergent structure in tetragonal crystals of OmpF porin. *Structure* **3** : 1051–1059.

Penel, S., Pebay-Peyroula, E., Rosenbusch, J., Rummel, G., Schirmer, T., and Timmins, P. A. (**1998**) Detergent binding in trigonal crystals of OmpF porin from Escherichia coli. *Biochimie* **80** : 543–551.

Perrakis, A., Harkiolaki, M., Wilson, K. S., and Lamzin, V. S. (**2001**) ARP/wARP and molecular replacement. *Acta Crystallographica Section D Biological Crystallography* **57** : 1445–1450.

Phillips, J. C. and Hodgson, K. O. (**1980**) The use of anomalous scattering effects to phase diffraction patterns from macromolecules. *Acta Crystallographica Section A* **36** : 856–864.

Pompidor, G., D'Aléo, A., Vicat, J., Toupet, L., Giraud, N., Kahn, R., and Maury, O. (**2008**) Protein crystallography through supramolecular interactions between a lanthanide complex and arginine. *Angewandte Chemie (International ed in English)* **47** : 3388–3391.

Pompidor, G., Girard, E., Maillard, A., Ramella-Pairin, S., Bersch, B., Kahn, R., and Covès, J. (**2009**) Biostructural analysis of the metal-sensor domain of CnrX from Cupriavidus metallidurans CH34. *Antonie van Leeuwenhoek* **96** : 141–148.

Pompidor, G., Maury, O., Vicat, J., and Kahn, R. (**2010**) A dipicolinate lanthanide complex for solving protein structures using anomalous diffraction. *Acta Crystallographica Section D Biological Crystallography* **66** : 762–769.

Pos, K. M. and Diederichs, K. (**2002**) Purification, crystallization and preliminary diffraction studies of AcrB, an inner-membrane multi-drug efflux protein. *Acta Crystallographica Section D Biological Crystallography* **58** : 1865–1867.

Pos, K. M., Schiefner, A., Seeger, M. A., and Diederichs, K. (**2004**) Crystallographic analysis of AcrB. *FEBS letters* **564** : 333–339.

Purdy, M. D., Ge, P., Chen, J., Selvin, P. R., and Wiener, M. C. (**2002**) Thiol-reactive lanthanide chelates for phasing protein X-ray diffraction data. *Acta Crystallographica Section D Biological Crystallography* **58** : 1111–1117.

Ramachandran, G. N. and Sasisekharan, V. (**1968**) Conformation of polypeptides and proteins. *Advances in protein chemistry* **23** : 283–438.

Ramagopal, U. A., Dauter, M., and Dauter, Z. (**2003**) Phasing on anomalous signal of sulfurs : what is the limit ? *Acta Crystallographica Section D Biological Crystallography* **59** : 1020–1027.

Rosenow, M. A., Huffman, H. A., Phail, M. E., and Wachter, R. M. (**2004**) The crystal structure of the Y66L variant of green fluorescent protein supports a cyclization-oxidation-dehydration mechanism for chromophore maturation. *Biochemistry* **43** : 4464–4472.

Rossmann, M. (**1972**) The molecular replacement method : a collection of papers on the use of non-crystallographic symmetry. Gordon and Breach.

Russo, S. and Baumann, U. (**2004**) Crystal structure of a dodecameric tetrahedral-shaped aminopeptidase. *The Journal of biological chemistry* **279** : 51275–51281.

Sanchez, J., Hoh, F., Strub, M., and Aumelas, A. (**2002**) Structure of the cathelicidin motif of protegrin-3 precursor : structural insights into the activation mechanism of an antimicrobial protein. *Structure* **10** : 1363–1370.

Schiltz, M., Fourme, R., and Prangé, T. (**2003**) Use of noble gases xenon and krypton as heavy atoms in protein structure determination. *Methods in enzymology* **374** : 83–119.

Schoehn, G., Vellieux, F. M. D., Asunción Durá, M., Receveur-Bréchot, V., Fabry, C. M. S., Ruigrok, R. W. H., Ebel, C., Roussel, A., and Franzetti, B. (**2006**) An archaeal peptidase assembles into two different quaternary structures : a tetrahedron and a giant octahedron. *The Journal of biological chemistry* **281** : 36327–36337.

Schrodinger, L. (**2010**) The PyMOL molecular graphics system. Schrödinger, LLC.

Schüttelkopf, A. W. and van Aalten, D. M. F. (**2004**) PRODRG : a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallographica Section D Biological Crystallography* **60** : 1355–1363.

Seeger, M. A., Schiefner, A., Eicher, T., Verrey, F., Diederichs, K., and Pos, K. M. (**2006**) Structural asymmetry of AcrB trimer suggests a peristaltic pump mechanism. *Science* **313** : 1295–1298.

Sheldrick, G. M. (**2008**) A short history of SHELX. *Acta crystallographica Section A, Foundations of crystallography* **64** : 112–122.

Silvaggi, N. R., Martin, L. J., Schwalbe, H., Imperiali, B., and Allen, K. N. (**2007**) Double-lanthanide-binding tags for macromolecular crystallographic structure determination. *Journal of the American Chemical Society* **129** : 7114–7120.

Sterpone, F., Bertonati, C., Briganti, G., and Melchionna, S. (**2010**) Water around thermophilic proteins : the role of charged and apolar atoms. *Journal of physics Condensed matter : an Institute of Physics journal* **22** : 284113.

Strub, M. P., Hoh, F., Sanchez, J. F., Strub, J. M., Böck, A., Aumelas, A., and Dumas, C. (**2003**) Selenomethionine and selenocysteine double labeling strategy for crystallographic phasing. *Structure* **11** : 1359–1367.

Taylor, G. L. (**2010**) Introduction to phasing. *Acta Crystallographica Section D Biological Crystallography* **66** : 325–338.

Thomas, A. S. and Elcock, A. H. (**2006**) Direct observation of salt effects on molecular interactions through explicit-solvent molecular dynamics simulations : differential effects on electrostatic and hydrophobic interactions and comparisons to Poisson-Boltzmann theory. *Journal of the American Chemical Society* **128** : 7796–7806.

Thorn, A. and Sheldrick, G. M. (**2011**) ANODE : anomalous and heavy-atom density calculation. *Journal Of Applied Crystallography* **44** : 1285–1287.

Törnroth-Horsefield, S., Gourdon, P., Horsefield, R., Brive, L., Yamamoto, N., Mori, H., Snijder, A., and Neutze, R. (**2007**) Crystal structure of AcrB in complex with a single transmembrane subunit reveals another twist. *Structure* **15** : 1663–1673.

Ursby, T., Weik, M., Fioravanti, E., Delarue, M., Goeldner, M., and Bourgeois, D. (**2002**) Cryophotolysis of caged compounds : a technique for trapping intermediate states in protein crystals. *Acta Crystallographica Section D Biological Crystallography* **58** : 607–614.

Usón, I. and Sheldrick, G. M. (**1999**) Advances in direct methods for protein crystallography. *Current opinion in structural biology* **9** : 643–648.

Veesler, D., Blangy, S., Cambillau, C., and Sciara, G. (**2008**) There is a baby in the bath water : AcrB contamination is a major problem in membrane-protein crystallization. *Acta crystallographica Section F* **64** : 880–885.

Walden, H. (**2010**) Selenium incorporation using recombinant techniques. *Acta Crystallographica Section D Biological Crystallography* **66** : 352–357.

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (**2009**) Jalview Version 2, a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25** : 1189–1191.

Weik, M. and Colletier, J. P. (**2010**) Temperature-dependent macromolecular X-ray crystallography. *Acta Crystallographica Section D Biological Crystallography* **66** : 437–446.

Weiss, M. S. (**2001**) Global indicators of X-ray data quality. *Journal Of Applied Crystallography* **34** : 130–135.

Wells, A. F. (**1984**) Structural Inorganic-Chemistry. Acta Cryst.

Wood, K., Frölich, A., Paciaroni, A., Moulin, M., Härtlein, M., Zaccai, G., Tobias, D. J., and Weik, M. (**2008**) Coincidence of Dynamical Transitions in a Soluble Protein and Its Hydration Water : Direct Measurements by Neutron Scattering and MD Simulations. *Journal of the American Chemical Society* **130** : 4586–4587.

Ying, W. (**2008**) NAD +/NADH and NADP +/NADPH in cellular functions and cell death : regulation and biological consequences. *Antioxidants & Redox Signaling* **10** : 179–206.

Yu, E. W., McDermott, G., Zgurskaya, H. I., Nikaido, H., and Koshland, D. E. (**2003**) Structural basis of multiple drug-binding capacity of the AcrB multidrug efflux pump. *Science* **300** : 976–980.

Zgurskaya, H. I. and Nikaido, H. (**2000**) Multidrug resistance mechanisms : drug efflux across two membranes. *Molecular Microbiology* **37** : 219–225.

Zulauf, H. (**1991**) Detergent phenomena in membrane protein crystallization. 53–72, CRC Press.

Zwart, P. H., Afonine, P. V., Grosse-Kunstleve, R. W., Hung, L.-W., Ioerger, T. R., McCoy, A. J., McKee, E., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., Storoni, L. C., Terwilliger, T. C., and Adams, P. D. (**2008**) Automated structure solution with the PHENIX suite. *Methods in molecular biology* **426** : 419–435.

Partie de la thèse dépourvue de science : longs mais sincères remerciements

Ces quatre années grenobloise ont été riches en émotions et en découvertes, scientifiques certes, mais aussi humaines, sportives, ou encore à propos de soi-même. Ma première déclaration d'impôt, mon premier loyer et ma première publication scientifique auront été grenobloise. Commencer dans la Recherche par autant de Découvertes n'est pas rien. Le petit bourgeois parisien protégé s'est retrouvé propulsé dans la vie active, tel un électron dans un synchrotron, et ce, pour son plus grand bonheur. Comme c'est allé vite... Si vite.

Grenoble. La Chamechaude en point de fuite de la rue Thiers, l'Archange, les hivers aux Sept Laux, les serveuses de La Bobine, la Dent de Crolles, le Hasard, le Moucherotte et ses Trois Pucelles, le Basse-Terre, le magistral Taillefer éternellement enneigé qui domine la rue des Martyrs, le parc Paul Mistral et sa tour Perret, Chamrousse qui porte si bien son nom les soirs cléments de printemps, le Family's, la terrasse universitaire EVE, les pelouses du campus Stendhal et ses décolletés estivaux, l'Ardoise, l'aiguillète de Quaix en Chartreuse visible du pont du tramway de l'avenue Foch, la place de Gordes et sa crêperie médiocre, les joggings à la Bastille pour faire le plein en gaz carbonique, ceux le long du Drac couleur d'émeraude, le marché de l'Estacade, la cascade de l'Oursière, la place Saint André que le soleil ne quitte qu'en dernier lieu, les matchs de rugby à *La Nat'*, les trop rares façades colorées de la place Grenette, contrastées par le bleuté du mont Saint-Eynard, le Cours Beriat et son tramway, le beau parc de la Villeneuve et son tumulte sordide, le Subway, les concerts gratuits au printemps, les toutes petites terrasses plus ou moins italiennes du quai Saint-Laurent, l'aviron sur le gris-bleu de l'Isère, les reflets du Trièves dans le lac du Monteynard...

Je me suis laissé happé par le "Y grenoblois", qu'il s'agisse de l'appellation géographique ou de la principale caractéristique du micro-accent local.

Mais lorsqu'on s'attache autant à un endroit, ce n'est pas uniquement en fonction du paysage ou de son travail mais surtout à cause de toutes les personnes qui le peuplent.

Alors avant que vous entriez en scène, sachez que si l'envie de partir ne m'est pas encore venue à l'heure où j'écris ces lignes, c'est en grande partie de votre faute ! Si pour des raisons sentimentales et professionnelles je ne m'en tiendrai pas à ce constat, il est bien réel, et voilà : il sera maintenant écrit ici en toute lettre. Vous avez tous contribué à rendre cette histoire grenobloise si... brulante et agréable.

Richard et Eric, mes deux compères de laboratoires, respectivement directeur de thèse et codirecteur de thèse. . . puis directeur de thèse. Je souhaite que vous ouvriez le bal. Le premier était un chercheur pluridisciplinaire, infatigable, rêveur, optimiste, bon vivant, humble et pédagogue, incarnation de l'époque d'une Recherche libre malheureusement révolue. Le second est un chercheur plus moderne : réaliste, plus pragmatique et dont le sens de l'observation n'a d'égal que sa perspicacité. Mélangez le tout au shaker (et non à la cuiller) et vous obtiendrez un des meilleurs encadrements qui soit pour faire une thèse : un trio où Richard fut "l'Empereur", Eric, le "Maître Jedi" et moi... J'étais le "Padawan". Merci à vous deux pour ces runs synchrotron riches en émotions diverses, pour ces petits déjeuners ESRF et surtout, pour m'avoir presque toujours considéré comme un chercheur à part entière et non pas comme un étudiant. On s'était dit une fois que pour faire de la bonne Science, il était nécessaire de bien rigoler. Alors, sans prétention aucune, je crois qu'on peut dire qu'on se sera globalement bien marré !

Eric, merci pour m'avoir redonné de l'énergie en baignant la cabine de contrôle de ta musique favorite lors des nuits synchrotron les plus difficiles. Réaliser des phasages aux "atomes lourds" sur fond de "heavy metal", c'est tout un concept. Un grand merci pour m'avoir fait découvrir le fabuleux monde des Studios Ghibli et du génial Miyazaki. Je te dois quelques moments d'évasion télévisuels bien mérités. Enfin, je te suis très reconnaissant de m'avoir consacré un certain nombre de weekends et de jours fériés pour corriger mon manuscrit lors de ma difficile et stressante période de rédaction.

Richard. . . Comme ce petit paragraphe le concernant va être délicat à écrire. Je vais encore être tiraillé entre l'envie d'être un peu trop sincère et la relative tempérance qu'impose des remerciements de thèse. De la même façon que j'ai été déboussolé, au début, entre le fait qu'il était mon directeur de thèse alors que nous devenions, j'ai la prétention de le croire, amis. Et encore, voilà un "prétention de croire" qui ressemble typiquement à une pincette due au contexte. Cependant, une chose est sure, une chose que personne ne contredira : nous étions fait pour nous entendre. Comme il le disait si bien à certains moments sur mon balcon, un verre de mon meilleur rhum à la main : "la vie est douce". Ce rhum, à force, il me l'a complètement sifflé et, symboliquement ou pas, je n'en ai plus jamais retrouvé. Alors, à son philosophe postulat, j'ajouterai ceci : ce lundi 3 octobre 2011, la mort fut rude. Malgré les élégantes et réconfortantes observations que j'ai pu entendre, comme le fait qu'il ait disparu trois jours après la fin de mon contrat de thèse ou qu'Eric fut son premier thésard sur le projet lanthanide et moi le dernier, je crois que j'aurai à jamais le regret de ne pas avoir pu fêter dignement avec lui notre réussite, tant professionnelle qu'humaine. On aurait eu tellement de temps après la thèse pour faire encore les bons-vivants. Je le remercie ici pour sa bienveillance, sa patience, sa Science qu'il prenait le temps de transmettre, son (peut-être trop grand) sens de la pondération, son intelligibilité, son sens de l'humour (bien que moins polyvalent que le mien concernant certains sujets), ses casse-têtes géométriques ou mathématiques (espiègle, il n'aurait sans doute pas résisté à vous faire remarquer que le numéro de cette page est divisible par onze...), ses pizzas chez Pierre et pour m'avoir présenté sa famille dont il aimait tant parler : Juliette, Myrtil, Damien et Nicolas que je salue affectueusement ici. Alors... Un grand merci à ce petit moustachu d'une telle grandeur d'âme.

Qui vient ensuite ? J'entrevois encore du monde ! Car je ne compte oublier personne.

Mon ami Michel Thepaut dit *Michou*. Ma canne, mon appui, mon étançon mais aussi le MacGyver du laboratoire. Soutenir quelqu'un de bout en bout et à ce point devrait mériter un financement en soi. Que ce soit pour m'avoir imprimé et relié tous mes exemplaires de thèse parce que je ne tenais plus debout, pour tes précieux conseils pour la soutenance ou pour m'avoir apporté un dimanche de rédaction au labo, en compagnie de ta si belle Marlène que je remercie aussi, une énorme boite de bonbon d'encouragement : merci pour tout ça... Et pour tout le reste. Une seule ombre au tableau cependant : quel dommage que tu sois avec Marlène ! Meow.

Laissez-moi vous présenter maintenant *"ma" Sarah* Ancelet, un bijou romantique petit format, brun et au teint clair, qui a tant d'énergie à revendre et à transmettre. Alors que tu étais dans les parages depuis le début, ce n'est qu'en fin de thèse que nous avons décidé de rattraper le retard ! Et là, je te remercie pour beaucoup de choses : pour ton dynamisme contagieux, pour ne pas m'avoir "facilement laissé tomber" sur "cette ligne qui s'estompe", pour m'avoir présenté la Bastille sous un autre jour et surtout, pour ton soutien.

Un grand merci à mon ami *Guénol'* Mainsant qui a eu la bonne idée de rejoindre aussi Grenoble pour une thèse ! Si on m'avait dit ? On peut continuer comme ça longtemps si tu veux, ça me va très bien. Étant donné qu'on sera encore potes, je te dis à dans 50 ans à Vancouver pour un petit café sur mon balcon ?

Guillaume Monin, pourquoi m'a tu lâchement abandonné ? Ces conversations de ménagères autour d'un bon thé, nos séquences cinéma en amoureux et tes recettes de crêpes et galettes bretonnes me manquent. J'suis ému.

Mon local préféré et *alter ego* Nicolas Martin ainsi qu'Agnès Thevenot, je vous remercie ici pour tout ce qui ce qui serait indécent de décrire ici. Vous serez toujours les bienvenus, c'est acquis, comme vous le savez déjà, même si il n'y aura pas toujours du chou. Je vous souhaite le meilleur pour la suite.

Je voudrai remercier mon mentor montagnard, celui qui connait mieux la distance entre le sommet de l'Obiou et celui du Bonnet de Calvin que le nom de la rue de son bar préféré : Julien Marcoux. Merci à toi et à Mamé pour m'avoir évadé du labo, certaines fins d'après-midi caniculaires d'aout, pour aller plonger dans le lac du Monteynard. Merci pour les nombreuses soirées dans la colloc' du bonheur.

Je remercie spécialement Juliette *Trepcu* pour tous les bons moments partagé, que ce soit en tant que compatriote de rédaction ou le reste du temps ! Tu vois, à l'heure où j'écris ces lignes, je t'aime ! On verra demain si ça continue ?

Je tiens aussi à remercier *Chabrule* (Eric Chabrol), Lydia Caro, David & Maisa Neves pour tous les bons délires laborantins et les bonnes journées ensoleillées en Haute-Savoie. Remerciement particulier à *Gallou* Goret, pour ses raisonnements singuliers et pour toutes nos conversations plus qu'instructives.

Un orage bleu électrique de mercis à la divine Corinne Vives pour avoir illuminé de sa légendaire beauté nos journées laborantines, devenant de fait, ensoleillées. Avoir la plus belle femme de l'IBS dans sa géographie amicale est une chance dont je me flatte chaque jour (ah mais si) ! Une avalanche de mercis pour Aurélien Deniaud qui m'a chaperonné à mon arrivée au laboratoire et pour l'ensemble de son recueil de blagues dont lui seul a le secret.

Pour leurs messages, leurs coups de fil, pour être venu à Grenoble, que ce soit en vacances ou pour ma soutenance, parce qu'ils se sont toujours rendus disponibles lors de ma présence à la capitale et surtout pour leur patience vis-à-vis de l'ours taciturne que j'ai souvent été lors de cette thèse, je remercie avec une infinie affection mes bons vieux potes de Paris : *Pilou* Brandon, *Fannynou* Jaulmes, *Lolo* Mendy, *Kiki*

Bertuletti, Laurence Feraud, *Matt'* Dartois, François Laur, *Didou* Weisenburger, Gaelle Schmidt, Laura et Alfredo Raimundo, *Lapin* Alloin, Martial Sankar, Fabio Valente, François et Linda Montenvert, *FX* Da Cunha, Gladys Potirath, *Vass'* Kalokyris, Sébastien Pontio, Jérôme Glay, *TomTom* Brunel, Tristan Wagner, Vincent Tanguy et Claire, Colin Guillemant, Antoine Movschin. Par votre indéfectible amitié vous avez toujours su m'apporter, et parfois malgré vous, moult bouffées d'air frais, réconfort et soutien. Il y a des amitiés qui sont faites pour durer...

Gratitude et insondable reconnaissances aux potes isérois (très rarement d'origine) qui ont contribué à rendre ma vie grenobloise si désopilante : *Gaby* Hellio aux grands yeux bleus, *Zouki* Lebrouc et *So'* Foret, *Alex'* Appolaire, biochimiste culinaire ou tout court et photographe de talent, *Les Perrault* (Matthieu du même nom et Aurore Laurendeau), *Alex'* Regnault, Anne Obermann et *La planous* (Thomas Planes), mon dépressif préféré *Guigui* Bacques, Antonin Kermen, Blandine Boue, *Mimi* Lafabrie, *Pinou*, Daphna Fenel, Francesca Mastropietro, Gildas Mainsant (les bonnes idées, c'est de famille : merci d'être aussi venu faire une thèse à Grenoble), mon ami allemand Holger Euchner, Stephane Chavand, Isabelle Gauvin, Clotilde Effantin, *Jeannot* Letort, Julien *Trepcu*, Lauris Chaumond, Marcio, *Lolotte* et Warody Lombardi (les bonnes idées, c'est finalement breton : merci d'être aussi venu s'installer à Grenoble), Antoine Licinio, Woltech, Thomas Cutuil mon biophysicien militant, *Les Rupin* (Matthieu et Sandy), Myriam de l'Ardoise, Nicolas Rihet et Laurene Huck, Pierre Bottelin, Pierre Boué, Romain Jolivet, Sivan, Yann Huon de Kermadec, Thierry, Max, Ahmad Sultan, *Anne-So'* Couvent, Arnaud et Laureen (mes premiers voisins de palier qui sont partis en nous donnant leurs snowboards). Je remercie tout particulièrement Marie-Charlotte qui s'est débrouillée pour me concocter un si délicieux et si savoureux pot de thèse. Un autre remerciement tout particulier adressé à ma voisine de palier, Sylvette Maniguet, qui s'est donné la peine de corriger mon résumé de thèse en anglais.

Un bien affectueux remerciement à mon autre invincible contrefort, ma famille, pour leur soutien, qu'il soit moral, culinaire et parfois (bien que j'ai tout fait pour que non) financier : mes parents, Isabelle et Philippe Talon, ma sœur, *Kikou*, mon beau-frère, Raynald Bachelet, et ma grand-mère, Mauricette Talon. Je profite de ce paragraphe qui fait référence à la famille et au soutien moral ou culinaire pour remercier aussi ma marraine Elizabeth Strich ainsi que ma "belle famille" : Michel Peyret, Joelle Guzzon et Monique Dugas. Je remercie aussi tout particulièrement les laragnais Guy et Renée Queyrel ainsi que les Chastel, pour leurs accueils dans le paysage provençal, toujours agréables, et le délicieux miel, carburant indispensable à un ours pour faire une bonne thèse.

Je remercie aussi mes précieux collaborateurs avec qui j'ai appris à faire de la bonne science : le Dr. Olivier Maury du Laboratoire de Chimie de l'Ecole Normale Supérieure de Lyon, Arnaud Gautier et Lionel Nauton de l'équipe SEESIB de l'Institut de Chimie de Clermont Ferrand, le Dr. Eric Condamine de la plateforme RMN-1D de l'Institut de Biologie Structurale Jean-Pierre Ebel de Grenoble, le Pr. Thierry Prange de Laboratoire de Cristallographie et RMN Biologiques de l'Université Paris V René Descartes de Paris, le Dr. Nathalie Colloc'h de l'équipe CERVOxy du Laboratoire ISTCT de Caen, le Dr. Cécile Breyton de l'équipe SSIMPA du groupe Membrane & Pathogène de l'Institut de Biologie Structurale Jean-Pierre Ebel de Grenoble, le Dr. Fabio Sterpone et Maria Kalimeri du Laboratoire de Biochimie Théorique de l'Institut de Biologie Physico-Chimique de Paris, le Dr. Philippe Oger du Laboratoire de Géologie de l'Ecole Normale Supérieure de Lyon ainsi que toute l'équipe de la ligne de lumière FIP-BM30A de l'ESRF : le Dr. Jean-Luc Ferrer, Michel Pirocchi et Jacques Joly.

Pour leur accueil, pour la bonne ambiance qu'ils savent mettre dans un labo et pour toute l'aide précieuse qu'ils m'ont quotidiennement apportée, que ce soit lors de discussions, parfois assez animées, ou de façon plus pratico-pratique, je remercie très chaleureusement et en premier lieu les membres de mon ancienne équipe LPM : Céline Julian-Binard (le labo s'écroulerait sans toi !), Antoine Picciocchi, Gavin Fox (merci pour m'avoir transmis des techniques ancestrales de cristallo et pour avoir partagé tes découvertes musicales avec l'amateur de "cheesy pop" que je suis). En second lieu et pour les mêmes raisons, je remercie les membres de mon groupe E.L.M.A : Dominique Madern (on n'est pas toujours sur la même longueur d'onde mais tes malate déshydrogénases sont tout de même de formidables bestioles), Mylène Ferruit, Frank Gabel, Frederic Vellieux (merci pour ton aide précieuse en cristallo), Pierre Vauclare, Eva Rosenbaum, Asun Dura, Louise Lassale et Bruno Franzetti, notre mélancolique et grand chef à tous. Là encore, nous avons fait ensemble de la belle Science. Même si nous n'étions pas dans la même équipe, je remercie amicalement et au même titre Jacques Coves et Antoine Maillard pour leur aide et leur soutien.

Je remercie les membres de mon jury de thèse pour avoir pris la peine de juger mon travail de recherche : le Dr. Daniel Picot et le Dr. Vincent Villeret en tant que rapporteurs, le Dr. Giuseppe Zaccai, Andrew Thompson et le Dr. Olivier Maury en tant que membres du jury.

Je tiens aussi à remercier toutes celles et ceux qui lors de conversations de couloir ou de réunions-machine-à-café improvisées m'ont transmis, entre deux légèretés, leur Science ou leur soutien ! Grâce à vous et à nos conversation, vous m'avez parfois donné un grand nombre d'informations scientifiques, et même, sans le savoir, de bons coups de pouce. J'enfonce une porte ouverte mais vous êtes la preuve même que négliger ces conversations, qui peuvent paraitre futiles de prime abord, serait une grave hérésie. Bien des idées, si petites soient elles, peuvent émerger de ce type de conversations. J'ai d'ailleurs appris bien après ma thèse que le monde de l'entreprise à un terme pour cela : "le réseau parallèle" considéré comme bien plus efficace que "l'officiel". En racontant cela je pense à : Jean-Michel Jault, Guy Schoehn, Philippe Carpentier, Eric Forest, Delphine Blot, Christophe Moreau, Bernhard Brutscher, Adrien Favier, Izabel Bérard, Nicole Thielens, Jérôme Dupuis, Viviana Job, Pierre Legrand, Bill Shepard et Roger Fourme.

Je remercie Blandine Cadiou, Chantal Robesson et Dominique Ribeiro pour leur accueil mais surtout pour avoir eu l'infinie délicatesse de corriger voire d'anticiper mes bévues administratives tout au long de cette thèse. Je remercie aussi mes geeks pour les divers dépannages informatiques inopinés et les moments de craquage délirants : Didier Depoisier, Jean-luc Parouty et surtout Fabrice Leger : une pinte nous attend toujours mon *Fab'*.

Il m'est impossible de ne pas remercier Felix Romain, mon encadrant de stage de seconde année de Master à l'Institut Pasteur, qui m'a mis le pied à l'étrier de la Recherche.

Je sais pertinemment que remercier un président d'Université peut paraitre louche ou gentiment obséquieux dans un contexte de sincères remerciements. Mais si le Professeur Farid Ouabdesselam, président de l'Université Joseph Fourier de Grenoble, ne m'avait pas accordé cette "bourse président", vous ne seriez pas en train de lire tout ceci.

Pour fermer la marche, je souhaite remercier, et le terme est trop faible, ma compagne, ma belle et douce Marine. D'un infaillible soutien, tu m'as supporté contre vents et marrées tout du long, que ce soit de près, heureux à Grenoble, ou de loin, depuis Dijon. Ta présence rassurante et apaisante a eu plus d'un effet salvateur. Merci pour avoir donc presque tout fêté avec moi, que ce soit les plus beaux résultats scientifique avec Richard, que tu appréciais tout autant, ou plus simplement, la vie.

Cette thèse concerne le développement de complexes de lanthanide et leur utilisation pour résoudre les structures de macromolécules biologiques par diffraction des rayons X, en particulier celles de protéines membranaires et de complexes protéiques de grande taille. Les complexes de lanthanide sont formés d'un atome de lanthanide et d'un ligand chimique qui assure une liaison non-covalente avec les surfaces protéiques. Introduits dans les cristaux de protéine, ces derniers constituent une sous-structure qui, déterminée par les méthodes de phasage *de novo* courantes, permet de résoudre la structure de la macromolécule d'intérêt.

La première partie de ce travail de thèse est une étude menée sur la grande famille des complexes picolinates de lanthanide, complexes luminescents dont la fixation au sein des cristaux est aisément décelable. En premier lieu, nous avons défini les conditions dans lesquelles le complexe tris-dipicolinate de lanthanide peut être utilisé ainsi que ses éventuelles capacités à promouvoir la cristallisation (effet supramoléculaire). En second lieu, un nouveau complexe, dérivé du précédent, a été développé au cours de cette thèse : le tris-hydroxymethyltriazoledipicolinate de lanthanide (« $[Ln(TDPA)_3]^{3-}$ »). Il nous a permis de réaliser un phasage *de novo* très précis conduisant à la détermination des structures du lysozyme de blanc d'œuf de poule et de la thaumatine de *Thaumatococcus daniellii* à haute résolution. Par ailleurs, différents ligands pour ce nouveau complexe ont été synthétisés par chimie-click, nous permettant de créer une panoplie de complexes uniques et des complexes hybrides. Nous avons montré que cette nouvelle famille de complexes présente une meilleure affinité pour les protéines permettant leur utilisation à de faibles concentrations. Les essais menés avec ces LnTDPA ont aussi permis d'entrevoir l'importance de la charge globale pour expliquer l'effet supramoléculaire. En troisième lieu, un tripicolinate cagé, le LnTNTPA, a également été évalué. Constitué d'une cage chimique de charge globale nulle, nous avons montré que ce nouveau complexe luminescent est le seul de cette famille picolinate qui puisse être utilisé en présence d'ions divalents.

Dans la seconde partie de cette thèse, nous décrivons l'utilisation des complexes de lanthanide pour le phasage de protéines multimériques de grandes tailles par les méthodes de phasage *de novo*. Premièrement, la structure de l'aminopeptidase dodécamérique PhTET1-12s de 480 kDa a été déterminée à 4 Å de résolution à l'aide du tris-dipicolinate d'europium. Ceci nous a permis de démontrer que les complexes de lanthanide peuvent être utilisés pour obtenir un phasage précis, même à basse résolution. Deuxièmement, les complexes de lanthanide issus de l'imagerie médicale ont aussi permis de déterminer la structure de trois nouvelles enzymes homotétramériques de la famille des malate déshydrogénases. Ces structures permettent d'apporter de nouveaux éclaircissements sur l'adaptation halophile. Enfin, en utilisant ces enzymes en tant que bibliothèque de fonctions chimiques, nous avons pu mettre en place une nouvelle approche méthodologique pour comprendre finement les modes d'interaction des complexes de lanthanide.

MOTS CLES : biologie structurale, cristallographie, méthodologie, diffraction aux rayons X, phasage *de novo*/expérimental, diffusion anomale, SAD, MAD, protéine multimérique, complexe de lanthanide.

This thesis aims to develop lanthanide complexes for solving biological macromolecular structures, especially membrane proteins and large protein complexes. Lanthanide complexes are composed of a lanthanide atom and a chemical ligand, which provides non-covalent binding to protein surfaces. Incorporated in protein crystals, they make up the substructure, determined by the widely-used *de novo* phasing methods, needed for solving the whole protein structure.

The first part of the present work shows a study of the luminescent lanthanide picolinate complexes family, easily detectable in protein crystals. First, we defined the conditions in which the known lanthanide tris-dipicolinate can be used and we examined its possible ability to promote protein crystallization (supramolecular effect). Secondly, a newly developed lanthanide tris hydroxymethyltriazoledipicolinate complex "$[Ln(TDPA)_3]^{3-}$", derived from this previous complex, allowed us to obtain a very precise *de novo* phasing for solving the high-resolution structure of the hen egg white lysozyme and the thaumatin from *Thaumatococcus daniellii*. Besides, various ligands for these new complexes were synthetized by click chemistry, enabling us to create both unique complexes outfit and hybrid complexes. We have shown that this new complexes family presents a better affinity for proteins allowing their use at very low concentrations. Tests conducted with those LnTDPA have also evidenced the importance of complex global charge to explain the supramolecular effect. Third, a caged tripicolinate, the LnTNTPA, was evaluated. Characterized by a neutrally charged chemical cage, we have shown that this new luminescent complex is the only one of the picolinate complexes family that can be used in the presence of divalent ions.

In the second part of this thesis, we describe the use of lanthanide complexes for phasing large multimeric proteins by *de novo* phasing methods. First, the structure of the 480 kDa dodecameric aminopeptidase PhTET1-12s was solved at 4 Å resolution by using the europium tris-dipicolinate which demonstrates that the lanthanide complexes can be used to obtain an accurate phasing, even at low resolution. Secondly, the lanthanide complexes from medical imaging also helped to solve the structures of three new homotetrameric enzymes from the malate dehydrogenases family. Those structures provide new insights into halophilic adaptation. Finally, by using these enzymes as a library of chemical functions, we developed a new methodological approach for a better understanding of the precise binding modes of those lanthanide complexes.

KEYWORDS: structural biology, crystallography, methodology, X-ray diffraction, *de novo*/experimental phasing, anomalous diffraction, SAD, MAD, multimeric protein, lanthanide complex.